ABSTRACT

Title of dissertation: EXTENDING THE LÉVY PROCESSES TO MULTIASSET PRODUCTS PRICING

Qing Xia, Doctor of Philosophy, 2006

Dissertation directed by: Professor Dilip B. Madan
Department of Finance

Lévy processes have gained great success in pricing single asset options. In this thesis, we introduce a methodology enabling us to extend the single asset pricing technique based on Lévy processes to multiasset cases.

In our method, we assume the log-return of each asset as a linear sum of independent factors. These factors are driven by the Lévy processes, and the specific Lévy process we are studying in this thesis is the Variance Gamma (VG) process. We recover these factors by a signal processing technique called independent component analysis (ICA), from which we get the physical measure ($\mathbb{P}$ measure). To price the contingent claims, we still need the risk-neutral measure ($\mathbb{Q}$ measure). We bridge the gap between physical measure and risk-neutral measure by introducing the transformation of measure between the $\mathbb{P}$ measure and the $\mathbb{Q}$ measure. We next write each asset as the linear sum of factors under risk-neutral measure. Thus, we may calibrate the measure change parameters simultaneously through individual listed option data. With the measure change parameters (from $\mathbb{P}$ measure to $\mathbb{Q}$ measure) being recovered, we're able to price multiasset products by doing Monte

Carlo simulation.

In this thesis, we also explore the possibility of extending Lévy processes to multiasset product pricing by applying the copula method. Generally speaking, the copula method enables us to introduce the dependence structure for arbitrary marginal distributions. The probabilistic interpretation of copulas is that we may apply the copula method to write the multivariate distributions for any marginal distributions. We consider examples from two different copula families - the elliptical copula family and Archimedean copula family. We studied Gaussian and Clayton one factor copulas as the examples from these two classes. We calibrated the correlation parameters for both methods and found them inconsistent across different strikes and maturities. And like the volatility smile and skew in the Black-Scholes model, we call it the skew and smile effect of correlation for one factor copula method.

EXTENDING THE LÉVY PROCESSES TO MULTIASSET PRODUCTS PRICING

by

Qing Xia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Dilip B. Madan, Chair/Advisor
Professor Michael Fu, Co-Advisor
Professor Jian-Guo Liu
Professor John E. Osborn
Professor Gurdip S. Bakshi

DEDICATION

To my mother and father

ACKNOWLEDGMENTS

First of all, I would like to thank all the people who inspired and encouraged me during my study here. Especially, I would like to thank my advisor Professor Dilip B. Madan, who opened the door of mathematical finance for me, whose insight guided me through all the time, who claimed himself to be a Professor plus a Quant. I am so glad to be able to share his sbrilliant ideas and always consider myself extremely lucky to have him as my advisor and mentor.

I would also like to thank Professor Michael Fu. His one year course sequence on stochastic process and simulation is still one of my favorites. I benefited a lot from his guidance and inspiration. I also want to thank Professor Jian-Guo Liu for his constant support and encouragement.

I would like to thank Professor John E. Osborn and Professor Gurdip S. Bakshi for agreeing to sit on my thesis committee and spending time reviewing my manuscript.

I want to especially thank the participants of Mathematical Finance RIT. This is a wonderful community which enriches my graduate life.

Finally, I want to thank my parents who always stand by me with their unbounded love. Words alone cannot express my gratitude.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In the case of single asset, the analysis of univariate time series data on financial asset shows that the log-returns are skewed (Kraus and Litzenberger [36] , Harvey and Siddique [30]) and have an actual kurtosis higher than that of the Normal distribution (Fama [22] , Cover and Thomas [17]). So, to better model the underlying asset, we need more flexible distributions. A number of papers proposed Lévy processes for describing the process of the log return (Eberlein, Keller and Prause [19], Madan, Carr and Chang [42], and Barndorff-Nielsen [8]). Carr, Geman, Madan and Yor [9] further explored the ability of infinite activity jump processes to capture both frequent small moves and rare large moves, which leads to the result that in the presence of an infinite activity jump processes one may dispense with the use of a diffusion component.

The successes with these processes showed that they're very capable of explaining the unconditional return densities on asset prices. We also see the successes with these processes for determining option prices by applying Fourier analysis (Bates [3], Scott [50]). Carr and Madan [16] further developed an approach to fully exploit the computational power of the Fast Fourier Transform (FFT). Recently, we saw the paper on the more efficient fractional FFT algorithm (FRFT) (Chourdakis [12]). For multiple maturities, extensions incorporating stochastic volatility are introduced by

Schoutens [51], Carr, Geman, Madan and Yor [10]. We therefore wish to extend the univariate valuation procedure to the multivariate case, which is the case of contingent claims written on a portfolio of assets.

For the case of multiasset, Markowitz [41] presented the traditional Mean-Variance preference investment analysis more than half a century ago. The Gaussian assumption simplifies the analysis but also incurs problems. The presence of skewness preference and kurtosis aversion is revealed by the analysis of single asset. But we know Gaussian models virtually ignore the higher moments, and it's investing on the assumption of zero information. So, we turn to the multidimensional counterpart of Lévy processes. The extension of the valuation technique to the high dimension case is not naive. The difficulties lie in the estimation of multidimensional probability laws and simulation of the multidimensional law.

Here we adopt the perspective of independent components analysis (ICA) (Hyvärinen, Karhunen and Oja [29]) introduced by Madan and Yen [46]. We postulate log-returns as a linear mixture of independent factors modelled by the Lévy processes widely used in the single asset literature. The specific Lévy process we are considering is Variance Gamma (VG) process. We apply the ICA technique to recover the mixing matrix and the independent factors. By doing so, we decompose the observed time series into statistically independent components (ICs) following the VG law. We thus reduce the multidimensional problems to a one-dimensional problem. The independent factors we get are actually governed by the statistical law (Physical Measure). To price the option prices, we still need the risk-neutral measure. Jacod and Shiryaev [34] show that the change of measure process can

be explicitly computed from the statistical and risk-neutral Lévy measures for pure jump processes. We construct the measure change process and calibrate the measure change parameters simultaneously for all the underlying assets from the listed option market data. The risk-neutral asset price process can then be discovered by the linear mixture of independent factors through the identified mixing matrix. With the measure change parameters being recovered, we're able to price the contingent claims written on multiasset by doing Monte Carlo simulation.

Another candidate of extending Lévy processes to multiasset product pricing is the copula method. The copula method was developed to study the dependence structure and construction of multivariate distributions. Sklar [55] shows that for any given continuous marginal distributions and multivariate distribution, the dependence structure can be separately described by a unique copula function. The copula method introduces the dependence structure for arbitrary marginal distributions to form a multivariate distribution. It builds the dependence upon the uniform random variables. A number of authors have proposed to apply copula method to price credit derivative product, Collateralized Debt Obligations, and other multiasset products (Li [38], Burtschell, Gregory and Laurent [5, 37], Madan, Konikov and Marinescu [43], Berd, Engle and Voronov [4], and Luciano and Schoutens [40]). Here, we focused on a one factor copula and studied a Gaussian copula from the elliptical copula family and select a Clayton copula (Clayton [14]) from the Archimedean copula family (Schweizer and Sklar [53] and Ling [39]). We form the multivariate distribution through a one factor copula function and take VG as the marginal distribution. We first calibrate the marginal distribution for each asset through listed

option data, then calibrate the copula parameters through our ICA basket option prices. We observed that for both copulas the parameters are not consistent with different strike prices and maturities. Like the volatility smile and skew in the Black-Scholes model, we call it the the smile and skew effect of parameters for one factor copula method.

Chapter 2

Preliminaries

The celebrated Black-Scholes model is succinctly expressed and beautifully formulated. But, we know that there are lots of problems with this model. One big problem with the model is that it assumes the log-return of the asset has the normal distribution and it treats volatility as a constant. If we naively apply the Black-Scholes pricing formula to the real options market, we could back out the implied volatilities for different strikes and observe the so-called volatility skew and smile [28]. Instead, we want a class of models to be able to capture the skewness and excess kurtosis features observed in the time series data of the asset log-return. Since late 80's, a number of papers proposed Lévy processes for modelling the log-return of stocks (Madan and Seneta [45], Barndorff-Nielsen [7] and Schoutens [51]). The success of modelling the log-return by Lévy processes for single asset motivates us to extend the methodology to multiasset case.

The main purpose of our study is to extend the Lévy process to multiasset products pricing. Thus, we'll start by reviewing Lévy processes. In this chapter, we'll briefly review the fundamental concepts of the Lévy process and its properties. The Variance Gamma process is thoroughly treated as a special example from Lévy processes. The Carr-Madan [16] FFT method and fractional FFT are also covered.

2.1 Lévy Processes: Definitions and Properties

In this section, we briefly review Lévy processes and its application to asset pricing. We'll go over the definition of Lévy processes, the Lévy-Khintchine formula and some important properties of Lévy processes.

2.1.1 Lévy Processes and Infinitely Divisible Distributions

Definition 2.1 *A stochastic process* $\{X_t; t \geq 0\}$ *on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$ *is a Lévy process if the following properties are satisfied:*

1. X_t *has independent increments:* $\forall\ n \geq 1$ *and* $0 \leq t_0 < t_1 < \cdots < t_n$*, the random variables* X_{t_0}, $X_{t_1} - X_{t_0}$, $\cdots$, $X_{t_n} - X_{t_{n-1}}$ *are independent.*

2. X_t *has stationary increments: the law of* $X_{t+h} - X_t$ *does not depend on* t*.*

3. $X_0 = 0$ *almost surely.*

4. X_t *is stochastically continuous:* $\forall\ t \geq 0$ *and* $\forall\ \epsilon > 0$*,*

$$\lim_{h \to 0} P\left(|X_{t+h} - X_t| > \epsilon\right) = 0.$$

5. X_t *is right-continuous with left limits almost surely.*

It's very hard for us to study the distribution of Lévy processes directly [34]. Instead, we study the Lévy processes through a powerful tool - the characteristic functions. The Lévy-Khintchine formula, established by Paul Lévy and A. Khintchine, gives the characterization of infinitely divisible random variables through their characteristic functions. This formula builds the one-to-one correspondence

between the Lévy process and its characteristic function. Moreover, the characteristic functions usually bear relatively simple forms, and it's easy for us to study them.

Definition 2.2 *A random variable X taking values in $\mathbb{R}^d$ is infinitely divisible, if for all $n \in \mathbb{N}$, there exists i.i.d random variables $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that*

$$X \stackrel{d}{=} Y_1^{(n)} + \cdots + Y_n^{(n)}. \tag{2.1}$$

Let μ be the law of X and $\phi_\mu(u)$ be the characteristic function of X, where $u \in \mathbb{R}^d$, then one may show the following:

Proposition 2.3 *The law μ of X is infinitely divisible if and only if for each $n \in \mathbb{N}$ there exists $\mu^{1/n}$ such that*

$$\phi_\mu(u) = \left[\phi_{\mu^{1/n}}(u)\right]^n. \tag{2.2}$$

2.1.2 Lévy-Khintchine Representation

The Lévy-Khintchine Representation builds the connection between infinitely divisible distributions and the Lévy processes. This formula enables us to study the Lévy processes through studying infinite divisible distributions, which have the representation in the form of characteristic functions. The representation is called Lévy-Khintchine formula.

Let μ be an infinitely divisible distribution of X on $\mathbb{R}$, and $\phi(u)$ be the characteristic function. The characteristic exponent is defined as $\psi(u) = \log \phi(u)$.

Theorem 2.4 (Lévy-Khintchine)

1. *Let μ be the infinitely divisible distribution on $\mathbb{R}$. Then the characteristic exponent $\psi(u) = \log \phi(u)$, where $\phi(u)$ is the characteristic function, is given by*

$$\psi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}\setminus\{0\}} (\exp(iux) - 1 - iux1_{|x|<1})\nu(dx), \tag{2.3}$$

where $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a measure on $\mathbb{R} \setminus \{0\}$ such that

$$\int_{\mathbb{R}\setminus\{0\}} (1 \wedge x^2)\nu(dx) < \infty. \tag{2.4}$$

In addition, the representation is unique.

2. *Conversely, if $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a measure satisfying (2.4), then there exists an infinitely divisible distribution whose characteristic exponent is given by (2.3).*

The (γ, σ, ν) in equation (2.3) is the generating triplet of the infinitely divisible distribution μ, and it's usually called the Lévy triplet. The measure μ is called the Lévy measure. If the Lévy measure is of the form $\nu(dx) = k(x)dx$, we call $k(x)$ the Lévy density. From (2.3), we observe that a Lévy process consists of three independent components: a deterministic drift, a continuous Brownian motion, and a pure jump process.

Now, let's discuss about the path properties of Lévy processes. If $\sigma^2 = 0$, the Lévy process is a pure jump process. If $\int_{|x|\leq 1} \nu(dx) < \infty$, the Lévy process is of *finite activity* (finite arrival rate of jumps); and if we have $\int_{|x|\leq 1} \nu(dx) = \infty$ instead, the Lévy process has *infinite activity* (infinite arrival rate of jumps).

If $\sigma^2 = 0$ and $\int_{|x|\leq 1} |x|\nu(dx) < \infty$, then the process is a pure jump process with finite variation. For this case, the characteristic exponent has the following simple form:

$$\psi(u) = i\gamma' u + \int_{\mathbb{R}\setminus\{0\}} (\exp(iux) - 1)\nu(dx), \tag{2.5}$$

where γ' is the new deterministic drift coefficient. A Lévy process with finite variation can be expressed as the difference of two increasing Lévy processes, which are also called *subordinators*.

2.1.3 Lévy-Itô Decomposition

Theorem 2.5 (Lévy-Itô decomposition) *Let X_t be a Lévy process. Then there exists $\gamma \in \mathbb{R}$, a Brownian motion $B_\sigma(t)$ with diffusion σ and an independent Poisson random measure N on $\mathbb{R}^+ \times (\mathbb{R} \setminus \{0\})$ such that, for each $t \geq 0$,*

$$X_t = \gamma t + B_\sigma(t) + \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} xN(t, dx), \tag{2.6}$$

where $\tilde{N}(t, x)$ is the compensated Poisson random measure given by

$$\tilde{N}(t, dx) = N(t, dx) - t\mu(dx),$$

where μ is the jump intensity given by $\mu(dx) = \mathbb{E}(N(1, dx))$.

The Lévy-Itô formula states that any Lévy process can be decomposed into two independent parts: the continuous part described by a Brownian motion and a pure jump part described by a Poisson process. And we could further decompose the jump terms into two parts: one part $\int_{|x|<1} x\tilde{N}(t, dx)$ describes the compensated sum of small jumps, the other part $\int_{|x|\geq 1} x\tilde{N}(t, dx)$ describes the 'large jumps'.

2.1.4 Measure Change for Lévy Processes

Measure change is an important tool in financial mathematics. It plays a key role in risk-neutral pricing, importance sampling and change of numeraire. A measure change really consists of reweighting the probability of paths. For a Brownian motion $B(t)$ under measure P, the classic Girsanov's theorem states how $B(t)$ will change under the new measure P' [48].

Theorem 2.6 (Girsanov) *Let B_t be a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_t$ be the accompanying filtration, and θ_u be a process adapted to $\mathcal{F}_t$. We define*

$$\begin{aligned} X_t &= B_t + \int_0^t \theta_s ds, \\ X_0 &= 0, \\ M_t &= \exp\left(- \int_0^t \theta_s dB_s - \frac{1}{2} \int_0^t \theta_s^2 ds \right). \end{aligned}$$

Also define a new measure P' equivalent to P, the Radon-Nikodym derivative is given by

$$\left. \frac{dP'}{dP} \right|_{\mathcal{F}_t} = M_t.$$

Assume that θ_u satisfies Novikov's Condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^T \theta_s^2 ds \right) \right] < \infty.$$

Then, X_t is a Brownian Motion under the new measure P'.

The Girsanov theorem states that if we change the drift of a given Brownian motion, then we can find an equivalent measure under which the new process is again

a Brownian motion. The law of the new Brownian motion will be absolutely continuous w.r.t. the law of the original Brownian motion, and we can compute explicitly the Radon-Nikodym derivative.

We have the similar results of measure changes between two Lévy processes [49, 18]. We may look the classic Girsanov's theorem as a special case of the following theorem.

Theorem 2.7 *Let X_t be a Lévy process on $\mathbb{R}$. Let P and P' be two measures specified by generating triplets (γ, σ, ν) and (γ', σ', ν') respectively. Then, $P|_{\mathcal{F}_t}$ and $P'|_{\mathcal{F}_t}$ are equivalent if and only if the following three conditions are satisfied:*

1. *They have the same diffusion component: $\sigma = \sigma'$,*

2. *The Lévy measures are equivalent with*

$$\int_{-\infty}^{\infty} \left(e^{\phi(x)/2} - 1\right)^2 \nu(dx) < \infty,$$

   where $\phi(x) = \ln\left(\frac{d\nu'}{d\nu}\right)$.

3. *If $\sigma = 0$, then we must in addition have*

$$\gamma' - \gamma = \int_{|x|\leq 1} x(\nu' - \nu)(dx).$$

Further, when P and P' are equivalent, the Radon-Nikodym derivative is

$$\left.\frac{dP'}{dP}\right|_{\mathcal{F}_t} = e^{U_t}$$

where

$$\begin{aligned} U_t &= \eta X_t^c - \frac{\eta^2\sigma^2 t}{2} - \eta\gamma t \\ &\quad + \lim_{\epsilon\downarrow 0}\left(\sum_{s\leq t,\ |\Delta X_s|>\epsilon} \phi(\Delta X_s) - t\int_{|x|>\epsilon}\left(e^{\phi(x)} - 1\right)\nu(dx)\right). \end{aligned} \quad (2.7)$$

Here, X_t^c is the continuous part of X_t, $\eta \in \mathbb{R}$ is chosen such that

$$\gamma' - \gamma - \int_{|x|\leq 1} x(\nu' - \nu)(dx) = \sigma^2\eta$$

if $\sigma > 0$, and $\eta = 0$ if $\sigma = 0$.

U_t is a Lévy process with generating triplet $(\gamma_U, \sigma_U, \nu_U)$ given by

$$\begin{aligned}
\sigma_U &= \sigma^2\eta^2, \\
\nu_U &= \nu\phi^{-1}, \\
\gamma_U &= -\frac{1}{2}\sigma^2\eta^2 - \int_{-\infty}^{\infty}(e^y - 1 - y1_{|y|\leq 1})(\nu\phi^{-1})(dy).
\end{aligned}$$

This result shows that for Lévy processes, we have much more freedom in changing measure while maintaining the equivalence of measures. But, if the diffusion component is absent, we cannot freely change the drift. Under this case, we may instead change the distribution of jumps $\nu(dx)$. We'll further discuss the change of measure for pure jump processes in Chapter 4.

2.1.5 Subordination of Lévy Processes

Let's now introduce another important concept *subordination*. Subordination enables us to transform from a Lévy process to a new Lévy process through random time change by an increasing Lévy process.

Definition 2.8 *A real-valued Lévy process is called a subordinator if it has nondecreasing paths almost surely.*

Lots of popular Lévy processes are generated by subordination. For example, the Variance Gamma process is a Brownian motion subordinated by a Gamma pro-

cess [45], the Normal Inverse Gaussian process is a Brownian motion subordinated by a Inverse Gaussian process [7], the Lévy stochastic volatility model is a Lévy process subordinated by the integration of a CIR process [10].

2.2 The Variance Gamma Process

The specific Lévy process employed in our study is the Variance Gamma process. The VG process is an example of pure jump process with infinite activity. The infinite activity property enables a pure jump process to capture both frequent small moves and rare large moves, which eliminates the need for diffusion component (Carr, Geman, Madan and Yor [9]. The VG process was originally introduced as a gamma time changed Brownian motion (Madan and Seneta [45]). Its Lévy density and characteristic function have very simple forms. It could also be interpreted as the difference of two independent Gamma processes (Madan, Carr and Chang [42]) by the fact that it's a process with finite variation.

2.2.1 Define the VG Process as Subordinated Brownian Motion

The VG process was originally derived by evaluating Brownian motion with drift at a random time given by a Gamma process. Let $b(t;\theta,\sigma)$ be a Brownian motion with constant drift θ and volatility σ :

$$b(t;\theta,\sigma) = \theta t + \sigma W(t), \tag{2.8}$$

where $W(t)$ is a standard Brownian motion.

Definition 2.9 *The Gamma process $\gamma(t;1,\nu)$ is a process with independent gamma increments of mean h and variance νh over non-overlapping time interval $(t, t+h)$. The independent increment $g = \gamma(t+h;1,\nu) - \gamma(t;1,\nu)$ follows the Gamma density $f_h(g)$:*

$$f_h(g) = \frac{g^{(\frac{h}{\nu}-1)}\exp(-\frac{g}{\nu})}{\nu^{\frac{h}{\nu}}\Gamma(\frac{h}{\nu})}, g > 0, \tag{2.9}$$

where $\Gamma(x)$ is a Gamma function. Its Lévy density is given by:

$$k_\gamma(x) = \frac{\exp(-x/\nu)}{\nu x}, x > 0, \tag{2.10}$$

and its characteristic function is given by:

$$\phi_\gamma(u) = \left(\frac{1}{1-iu\nu}\right)^{h/\nu}. \tag{2.11}$$

Note that $\int_0^\infty k_\gamma(x)dx = \infty$ and $\int_0^\infty x k_\gamma(x)dx < \infty$, so the Gamma process has infinite activities and finite variation. The process is a pure jump process and may be approximated as a compound poisson process.

Having defined the Gamma process, we can define the Variance Gamma process.

Definition 2.10 *Let $b(t;\theta,\sigma)$ be a Brownian motion defined in Equation (2.8) and $\gamma(t;1,\nu)$ be a Gamma process defined in Definition 2.9, then the VG process $X(t;\sigma,\nu,\theta)$ is defined as*

$$X(t;\sigma,\nu,\theta) = b(\gamma(t;1,\nu);\sigma,\theta). \tag{2.12}$$

We observe that the VG process $X(t)$ is a Brownian motion evaluated at a random time given by a Gamma process. Assuming the market evolves as a gamma

time changed Brownian motion has its economic explanation. The market does not evolve equally all the time, sometimes there're lots of trading activities, while for other times there are less trading activities. Thus, the length of a day doesn't follow the calendar time, instead, it follows a random " business time " - following a Gamma distribution for our case.

2.2.2 Properties of the VG Process

The VG process has lots of nice properties. Its characteristic function and Lévy density have very simple forms. It's a process with infinite activity and finite variation. We'll explore these properties in this section.

We know that the VG process is a gamma time changed Brownian motion:

$$X(t;\sigma,\nu,\theta) = b(\gamma(t;1,\nu);\sigma,\theta) = \theta\gamma(t;1,\nu) + \sigma W(\gamma(t;1,\nu)),$$

So, the characteristic function of the VG process can be obtained by conditioning on the gamma time and noticing that conditioned random variable is Gaussian. Then, we apply the Laplace transform to get the unconditioned characteristic function [42]:

$$\phi_{X(t)}(u;\sigma,\nu,\theta) = E(e^{iuX(t)}) = \left(\frac{1}{1 - i\theta\nu u + \sigma^2 u^2 \nu/2}\right)^{t/\nu}. \tag{2.13}$$

The characteristic function is infinitely divisible and Madan *et al.* [42] also showed that the VG process can be expressed as the difference of two independent Gamma processes. This can be done by showing that the characteristic function of VG process can be written as the product of two characteristic functions, and further notice that they are the characteristic functions of Gamma processes.

The Lévy density $k_{VG}(x)$ of VG process is identified as follows:

$$K_{VG}(x) = \begin{cases} C\exp(Gx)/|x|, & x < 0 \\ C\exp(-Mx)/x, & x > 0 \end{cases} \tag{2.14}$$

where

$$\begin{aligned} C &= 1/\nu > 0, \\ G &= \left(\sqrt{\frac{\theta^2\nu^2}{4} + \frac{\sigma^2\nu}{2}} - \frac{\theta\nu}{2}\right)^{-1} > 0, \\ M &= \left(\sqrt{\frac{\theta^2\nu^2}{4} + \frac{\sigma^2\nu}{2}} + \frac{\theta\nu}{2}\right)^{-1} > 0. \end{aligned}$$

With these parameters, one can write VG process as the difference of two independent Gamma processes:

$$X(t) = \gamma^+(t) - \gamma^-(t) \tag{2.15}$$

where $\gamma^+(t)$ is a Gamma process with parameters C and M (*with mean CM and variance CM^2*) and $\gamma^-(t)$ is a Gamma process with parameters C and G (*with mean CG and variance CG^2*).

In terms of C, G, M, we could also rewrite the characteristic function as follows:

$$\phi_{X(t)}(u; C, G, M) = \left(\frac{GM}{GM + iu(M-G) + u^2}\right)^{Ct}. \tag{2.16}$$

One observes that the integral of Lévy density around zero is infinite, which leads to the infinite activity property. The finite variation property is observed by noting $\int_{|x|<1} |x| k_X(x) < \infty$ and the fact that VG process has no Brownian component.

1. Generate a gamma r.v. $g \sim \text{Gamma}(t/\nu, \nu)$,
2. Generate a standard normal r.v. $z \sim \text{Normal}(0, 1)$,
3. The VG process at time t is given by : $X(t; \sigma, \nu, \theta) = \theta g + \sigma\sqrt{g}z$.

Table 2.1: Simulating the VG process $X(t; \sigma, \nu, \theta)$ as a Time Changed Brownian Motion

The VG process gains the control over both skewness and kurtosis. When $\theta = 0$, Lévy density is symmetric and the skewness is zero. Negative values of θ generate negative skewness. The $\nu = (1/C)$ provides the control over kurtosis.

2.2.3 Simulating the VG Process

The simulation of VG process is a direct application of it definition. We know that the VG process is defined as a Gamma time changed Brownian motion:

$$X_{VG}(t) = \theta\gamma(t) + \sigma W(\gamma(t)),$$

where $\gamma(t)$ is a Gamma process defined in Definition 2.9, and $W(t)$ is a Brownian motion. As a result, the VG process can be simulated as a Brownian motion sampled by a random time given by Gamma random variate. We present the algorithm of simulating the VG process as a time changed Brownian motion in Table 2.1.

We also learn from Equation (2.15) that a VG process can be written as the difference of two Gamma processes. Using the (C, G, M) parametrization, the VG process $X(t; C, G, M)$ is given by:

$$X(t; C, G, M) = \gamma^+(tC, 1/M) - \gamma^-(tC, 1/G), \tag{2.17}$$

1. Generate a Gamma process $\gamma^+(tC, 1/M)$,
2. Generate a Gamma process $\gamma^-(tC, 1/G)$,
3. The VG process at time t: $X(t; C, G, M) = \gamma^+(tC, 1/M) - \gamma^-(tC, 1/G)$.

Table 2.2: Simulating the VG process $X(t; C, G, M)$ as the Difference of Two Gamma Processes

where $\gamma^+(tC, 1/M)$ and $\gamma^-(tC, 1/G)$ are Gamma processes. We present the algorithm of simulating the VG process as the difference of two Gamma processes in Table 2.2.

2.2.4 The VG Stock Price Process

In this section, we introduce the risk neutral dynamics of the stock price modelled by the VG process.

The VG stock price dynamics is obtained by replacing the Brownian motion in the Black-Scholes geometric Brownian motion model by the VG process, and adding a convexity correction factor to make the discounted stock price a martingale. The risk neutral stock price process is given by:

$$S(t) = S(0)\exp(rt + X(t; \sigma, \nu, \theta) + \omega t), \tag{2.18}$$

where r is the continuously compounded interest rate, $X(t; \sigma, \nu, \theta)$ is the risk neutral VG process with parameters (σ, ν, θ), and let's take a quick review at how ω is determined. For the risk neutral VG stock process $S(t)$, we have:

$$S(t) = S(0)\frac{\exp(rt + X_{VG}(t))}{\mathbb{E}\left(\exp(X_{VG}(t))\right)}. \tag{2.19}$$

We want to make sure that the stock prices discounted by money market account are martingales, which means that we will choose ω so that $\mathbb{E}\left(S(t)\exp(-rt)\right) = S(0)$. By applying (2.13) we have

$$\mathbb{E}\left(\exp(X_{VG}(t))\right) = \phi_{VG}(-i) = (1-\theta\nu-\frac{1}{2}\sigma^2\nu)^{-t/\nu} = \exp(-\omega t), \tag{2.20}$$

then, by taking logarithm on both sides we'll get $\omega = \frac{1}{\nu}\ln(1-\theta\nu-\frac{1}{2}\sigma^2\nu)$.

Just like the way we derive the characteristic function for VG process, we can also get the density of the log stock price relative and express it in terms of the modified Bessel functions of the second type [42].

Theorem 2.11 *The density for the log price relative $z = \ln(S(t)/S(0))$ when prices follow the risk neutral VG stock process (2.19) is given by:*

$$g(z) = \frac{2\exp(\theta x/\sigma^2)}{\nu^{\frac{t}{\nu}}\sigma\sqrt{2\pi}\Gamma(\frac{t}{\nu})}\left(\frac{x^2}{2\sigma^2/\nu+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} \times K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{x^2(2\sigma^2/\nu+\theta^2)}\right), \tag{2.21}$$

where K is the modified Bessel function of the second type,

$$x = z - rt - \frac{t}{\nu}\ln(1-\theta\nu-\sigma^2\nu/2),$$

and the VG parameters are risk neutral ones.

Madan, Carr, and Chang [42] also give the characteristic function for the log of S_t:

$$\phi_{\ln(S_t)}(u) = \exp\left[iu(\ln(S_0)+(r+\omega)t)\right](1-i\theta\nu u+\frac{1}{2}\sigma^2u^2\nu)^{-t/\nu}, \tag{2.22}$$

and this equation is used to calculate the VG option prices by Fast Fourier Transform (FFT) method.

2.3 The VG Option Prices and FFT method

For VG process, we also have the close-form option prices. Madan *et al* [42] evaluate the VG option by first conditioning on the gamma random time change g. Noticing that conditional VG distribution $X|g$ has normal distribution, one may apply the Black-Scholes formula to get the conditional option price. They derived the analytic VG option price by integrating out the gamma variate. Carr and Madan [16] further developed an approach to fully exploit the computational power of the Fast Fourier Transform (FFT) method.

2.3.1 The Carr-Madan FFT Method

The Carr-Madan FFT method evaluate the value of an option by doing an inverse Fourier transform to the characteristic function of the log price. The method handles the the singularity in the integrand by including a dampening factor. The method is faster than using the analytic formula, which requires you to do a numerical integration of the modified Bessel function of the second type. We sketch the method as follows.

Let k be the log of the strike price K, and let $C_T(k)$ be the value of a call option with maturity T and strike $\exp(k)$. Let the $\phi_T(u)$ be the characteristic function of the risk neutral density of the log price S_T. Then, the call prices can be expressed as follows:

$$C_T(k) = \frac{\exp(-\alpha k)}{\pi} \int_0^\infty e^{-ivk} \psi_T(v) dv, \tag{2.23}$$

where $\exp(-\alpha k)$ is the dampening factor, $\psi_T(v)$ is defined as follows:

$$\psi_T(v) = \frac{e^{-rT}\phi_T\left(v-(\alpha+1)i\right)}{\alpha^2+\alpha-v^2+iv(2\alpha+1)}. \tag{2.24}$$

We may approximate the equation (2.23) using an integration rule, such as trapezoidal rule, and write it approximately:

$$C_T(k) \approx \frac{\exp(-\alpha k)}{\pi}\sum_{j=1}^{N} e^{-iv_j k}\psi_T(v_j)\eta, \tag{2.25}$$

where N is chosen to be the power of 2, η is the step size for the support of the characteristic function u, $\eta = a/N$, a is the upper limit for the integration (2.23), $v_j = (j-1)\eta$. If we choose the step size of the log strike k to be λ, then the log strikes change from $-b$ to b, so $k = -b + \lambda(u-1)$, for $u = 1, \cdots, N$, we have the following:

$$C_T(k) \approx \frac{\exp(-\alpha k)}{\pi}\sum_{j=1}^{N} e^{-i\lambda\eta(j-1)(k-1)}e^{ibv_j}\psi_T(v_j)\eta. \tag{2.26}$$

By further requiring $\lambda\eta = \frac{2\pi}{N}$ and applying Simpson's rule, we may write the call price as:

$$C(k) = \frac{\exp(-\alpha k)}{\pi}\sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)}e^{ibv_j}\psi(v_j)\frac{\eta}{3}(3+(-1)^j-\delta_{j-1}), \tag{2.27}$$

where δ_n is the Kronecker delta function. Let

$$x(j) = e^{ibv_j}\psi(v_j)\frac{\eta}{3}(3+(-1)^j-\delta_{j-1}),$$

then, we may further rewrite the equation (2.27) as follows:

$$C(k) = \frac{\exp(-\alpha k)}{\pi}\sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)}x(j), \tag{2.28}$$

Recall that FFT is an algorithm for computing the sum

$$w(k) = \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)} x(j), k = 1, \cdots, N, \tag{2.29}$$

we may apply FFT to equation (2.28) to compute the call prices.

The Carr-Madan FFT method is very efficient for evaluating the option prices. For one single run, it calculates the option prices across all the strikes. Another benefit for us to adopt this method is that it evaluates the call prices through the characteristic function. The analytic forms of characteristic functions for the Lévy processes like VG, NIG, Meixner, and CGMY are all known. So, we only need to change the characteristic functions for different models to put into the FFT "engine" to get the call option prices. This method also works for stochastic volatility Lévy models [10] and scaled self decomposable models [11].

2.3.2 The Fractional FFT Method

Recently, we saw the paper on the more efficient fractional FFT algorithm (FRFT) (Chourdakis [12]). Recall that to apply FFT method to option valuation in Carr and Madan [16], we need the restriction on the step size of log strikes and the support of the characteristic function as follows:

$$\lambda\eta = \frac{2\pi}{N}. \tag{2.30}$$

We observe that for the three parameters λ, η, N, there're only two can be arbitrarily chosen. Since the integration upper bound a is defined as:

$$a = N\eta, \tag{2.31}$$

1.	$\boldsymbol{y}_1 = \left((x_{j-1}e^{-i\pi(j-1)^2\beta})_{j=1}^N, (0)_{j=1}^N\right),$	$\tilde{\boldsymbol{y}}_1 = FFT(\boldsymbol{y}_1),$
2.	$\boldsymbol{y}_2 = \left((e^{i\pi(j-1)^2\beta})_{j=1}^N, (e^{i\pi(N-j+1)^2\beta})_{j=1}^N\right),$	$\tilde{\boldsymbol{y}}_2 = FFT(\boldsymbol{y}_2),$
3.	$\boldsymbol{y}_3 = \tilde{\boldsymbol{y}}_1 \odot \tilde{\boldsymbol{y}}_2,$	$\tilde{\boldsymbol{y}}_3 = IFFT(\boldsymbol{y}_3),$
4.	$D_k(\boldsymbol{x}, \beta) = \tilde{\boldsymbol{y}}_3 \odot (e^{-i\pi(k-1)^2\beta})_{k=1}^N.$	
	where $\odot$ denotes element-by-element vector multiplication	

Table 2.3: The FRFA Algorithm

λ is inversely proportional to the a. So, to make the grid spacing dense enough for the option pricing, which means that we choose the λ to be small enough, we have to set the a to be very big. As a result of that, Kyriakos [12] mentions that out of 4096 calculated option prices, only about 67 will fall within the log strike interval of practical applications. So, this requirement will lower the efficiency of the method.

For fractional FFT method, this restriction can be dropped. The fractional FFT (FRFT) is an efficient algorithm to compute the sum

$$w(k) = \sum_{j=1}^{N} e^{-i2\pi(j-1)(k-1)\beta} x(j), k = 1, \cdots, N,$$

and we may treat FFT as a special case of FRFT for $\beta = \frac{1}{N}$. Thus, we only need to have the restriction $\lambda\eta = 2\pi$ to apply the FRFT method. This means that the grid size of characteristic function support and log-strike can be chosen independently.

Generally, N-point FRFT can be implemented by invoking three 2N-point FFT. For a N-point FRFT procedure on $\boldsymbol{x} = (x)_{j=1}^N$, $D_k(\boldsymbol{x}, \beta)$, we present the FRFT algorithm in the Table 2.3.

As we may see in the Table 2.3, to do one FRFT, we have to do 3 FFT (with

fewer points). But, overall 256-point FRFT has similar elementary operations as a 1024-point FFT, with the same effect as 4096-point FFT [12]. Thus, FRFT is almost 4 times faster than FFT. This efficiency improvement will benefit us a lot when it comes to calculating the multiasset option.

Chapter 3

Independent Component Analysis

The independent component analysis (ICA) plays a key role in our proposed method. To extend the Lévy process to multiasset product pricing, we adopt the perspective introduced by Madan and Yen [46] recently. We postulate the log return for each asset as a linear mixture of factors driven by Lévy processes. These factors are unknown and independent of each other. To uncover these factors, we need to apply the technique called the independent component analysis (ICA).

Independent component analysis is a technique for uncovering hidden factors driving the observed multivariate data. To apply the ICA technique, we assume that the original multivariate data are the linear transformation of these unknown factors. The factors are assumed to be statistically independent and nongaussian. Other well-known linear transformation methods include principal component analysis and factor analysis. The detail treatment of linear transformation methods can be found in [33], [29], and [26]. This chapter follows the structure of [29].

3.1 Linear Transformation Methods

An important problem in disciplines such as statistics, signal processing and data analysis, is finding a suitable representation of the multivariate data, by means of a transformation. For computational and conceptual simplicity, the representa-

tion is often sought as a linear transformation [26].

3.1.1 The Linear Representations

Let X be an n-dimensional random variable, S is the n-dimensional unobservable random variable, then the linear transformation states that:

$$X = AS, \tag{3.1}$$

where A is the linear mixture matrix to be determined, X is observable and we want to find out the unobservable source S. The problem boils down to finding the hidden driving factors and the mixing matrix. This problem is also called the *cocktail-party* problem, in the sense that we may only observe the mixed drink while the original ingredients is unknown to us.

In fact, if we know the mixture matrix A, then we could easily recover the unknown source S. But, the problem is that we only know the LHS of the equation (3.1), the observable data X. So, we need to define the principles that find the suitable linear transformation. Some principles and methods have been proposed. Among these are principal component analysis, factor analysis, and independent component analysis.

The well-known principal component analysis (PCA) and factor analysis (FA) both belong to the second order methods. The second order methods find the representation using only the information contained in the the mean and the covariance matrix of the observed data. These methods also assume that the observed data X are normally distributed , and thus, the higher moments are ignored. Namely, these

methods only exploit the information coming from the first two moments, mean and variance. For simplicity, we assume that the variable X is centered, which means that $X = X_0 - \mathbb{E}(X_0)$.

3.1.2 Principal Component Analysis

Principal component analysis (PCA) is a powerful tool to analyze high dimensional correlated data. The idea of PCA is to reduce dimensions so that only the most important sources of information are kept. It achieves this goal by the decorrelation method. The method identifies a smaller set of underlying variables with less redundancy, which is measured by correlations between data elements. These variables are called principal components and they carry the most important information of the original data. By assumption, the principal components follow normal distribution. Thus, their independence with each other can be inferred from their uncorrelatedness.

Before applying PCA, we preprocess the data by subtracting it by its mean and normalizing its variance to be unit. Let X be the data with zero mean and unit variance, C be the covariance matrix of X. The goal of PCA is to find a linear transform represented by a orthogonal matrix $W : P = XW$, such that $cov(P)$ is a diagonal matrix, with diagonal entries decreasingly ordered by their value.

PCA applies eigenvalue and eigenvector analysis to C. Each principal component is a linear combination of the columns x_i of X. And the weights are chosen such that the first principal component explains the greatest amount of the total

variation of X, the second component explains the second greatest amount of the variation, and so on. PCA finds the $n \times n$ orthogonal matrix W:

$$C = W\Lambda W', \tag{3.2}$$

where Λ is the $n \times n$ diagonal matrix of eigenvalues of C. The columns of W are ordered according to the size for the corresponding eigenvalue so that the ith column of W, denoted by $w_i = (w_{1i}, \cdots, w_{ni})'$, is the $n \times 1$ eigenvector corresponding to the eigenvalue λ_i, where the eigenvalues have been ordered so that $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. Then, the ith principal component is given by:

$$P_i = w_{1i}x_1 + w_{2i}x_2 + \cdots + w_{in}x_n,$$

where $i = 1, \cdots, n$. Or in matrix notation,

$$P = XW. \tag{3.3}$$

To see that the principal components are uncorrelated, we apply equation (3.2) and the fact that W is an orthogonal matrix:

$$\mathbb{E}(P'P) = W'\mathbb{E}(X'X)W = W'CW = (W'W)\Lambda(W'W) = \Lambda,$$

Since Λ is a diagonal matrix, the columns of P are uncorrelated, and the variance of the ith principal component is λ_i and they are decreasing in value.

Recall again that W is an orthogonal matrix, $W' = W^{-1}$, then equation (3.3) can be rewritten as:

$$X = PW', \tag{3.4}$$

so, for original factor x_i, we have:

$$x_i = w_{i1}P_1 + w_{i2}P_2 + \cdots + w_{in}P_n, \tag{3.5}$$

equation (3.5) is the *principal components representation* of the original factors. Thus, the original data X are explained by n principal components and the dimension of the original system is reduced.

3.1.3 Factor Analysis

Another second order method closely related to PCA is factor analysis. The general model for the data is postulated:

$$X = AS + n \tag{3.6}$$

where X is the observed data, S is the vector of latent factors that cannot be observed, A is a constant matrix and the elements of A are called *factor loadings*, and n is noise vector of the same dimension as X. Both S and n are assumed to be Gaussian, zero-mean and white. In addition, the dimension of S is assumed to be lower than that of X. Thus, we can also view factor analysis as a dimension reduction method.

Factor analysis is intrinsically the modification of PCA. If we assume the covariance matrix $\Sigma = \mathbb{E}(nn')$ of the noise n is known, then the factors are found by performing PCA using the modified covariance matrix $C - \Sigma$, where C is the covariance matrix of X. Thus the vector S is the vector of principal components of X without noise. The problem (3.6) does not have unique solution. One must impose extra constraints to get a unique solution.

3.2 Definition of Independent Component Analysis

Both PCA and FA are second order method. The latent factors are assumed to be Gaussian, only mean and covariance of the observed data are used in the estimation. The factors are also assumed to be uncorrelated, then by the Gaussian assumption, they're independent. The independent component analysis (ICA) is a similar latent factor model, but the factors are assumed to be *statistically independent* and *nongaussian*. By removing the Gaussian assumption, ICA actually takes the higher order moments into account. The definition of ICA are given below:

Definition 3.1 *ICA of a random vector X consists of estimating the following generative model for the data:*

$$X = AS, \tag{3.7}$$

where the latent factor s_i in the vector $S = (s_1, \cdots, s_n)'$ are assumed statistically independent and nongaussian, the matrix A is an unknown mixing matrix.

For simplicity, we assume the number of the factors (independent components) is the same as the number of the observed data, which means A is an $n \times n$ square matrix.

Note that there're other definitions of ICA other than the definition (3.1). One could add noise term to model (3.7) like the factor analysis, one may also assume the number of the factors is not equal to that of the original data. The detailed discussion of these topics may be found in [29].

3.3 Principles in Estimating ICA

In the ICA mixing model (3.7), we need to uncover the hidden factors S on the right hand side. Since both A and S are unknown, we need to apply some principles on the hidden factors to find the factors with desirable properties. And we'll explore these restrictions and assumptions imposed in Definition (3.1) in this section.

3.3.1 Statistically Independent

The basic assumption of ICA is that the components s_i are statistically independent. Let's consider two random variables y_1 and y_2. The variables y_1 and y_2 are independent if the knowledge of the value of y_1 does not give any knowledge on the value of y_2, and vice versa. We could define independence by the probability densities function(pdf). Let's denote by $p(y_1, y_2, \cdots, y_n)$ the joint probability density function of the $(y_1, y_2, \cdots, y_n)$, and by $p_i(y_i)$ the marginal probability density function of y_i. Then the y_i are independent if and only if the joint pdf can be written as follow:

$$p(y_1, y_2, \cdots, y_n) = p(y_1)p(y_2)\cdots p(y_n). \tag{3.8}$$

Independence among random variables is a stronger requirement than lack of correlation. Independence implies zero covariance, but zero covariance does not always imply independence of random variables. For PCA and FA, the hidden factors are assumed to be Gaussian. We know that for Gaussian random variables, independent and uncorrelated are equivalent. In estimating PCA and FA, it is enough to require the factors be uncorrelated. But, for ICA, lack of correlation is not enough, and we

need the second assumption that the components follow nongaussian distribution.

3.3.2 Nongaussian Distribution

The next assumption of ICA is that the independent components must have nongaussian distributions. This is a very important requirement. In fact, the mixing matrix can not be inferred from the observed data, if the components have the Gaussian distribution. Let's take a look at an example. Assume that the joint distribution of two independent components s_1 and s_2 is Gaussian, then we have their joint pdf given by:

$$p(s_1, s_2) = \frac{1}{2\pi} \exp(-\frac{s_1^2 + s_2^2}{2}) = \frac{1}{2\pi} \exp(-\frac{\|S\|^2}{2}).$$

Also, we assume that the mixing matrix A is orthogonal. Since $X = AS$, the joint pdf of the observed data x_1 and x_2 is given by

$$p(x_1, x_2) = \frac{1}{2\pi} \exp(-\frac{\|A'X\|^2}{2})|\det A'|,$$

Notice that A is an orthogonality matrix, we have $\|A'X\|^2 = \|X\|^2$ and $|\det A| = |\det A'| = 1$. So, we have

$$p(x_1, x_2) = \frac{1}{2\pi} \exp(-\frac{\|X\|^2}{2}),$$

which means that the independent factors and the original data have the same distribution, and the mixing matrix does not change the joint pdf at all. This fact tells us that the mixing matrix can not be estimated with the Gaussian assumption. In fact, the ICA model can not be estimated if $s_1, s_2, \cdots, s_n$ are all Gaussian [29]. For Gaussian distribution, the higher order moments are ignored, and the information

from the first two moments is not enough to uncover the hidden factors. ICA is a process being able to utilize the information of higher order moments.

3.3.3 Unit Variance

One more assumption will be the variance of unknown source S. Recall in the ICA model $X = AS$, both S and A are unknown. If we multiple the source S by a scalar α, it could be cancelled out by dividing the matrix by the same scalar α:

$$X = (A\frac{1}{\alpha})(\alpha S).$$

In other word, we may have infinite number of solutions. Thus, we need to fix the magnitudes of the independent components s_i. To do so, we require that they all have unit variance: $\mathbb{E}(SS') = I$. With the magnitudes of the independent components being fixed, we could find the matrix A adapted to this condition.

3.4 Estimating the ICA Model

With the principles being defined, we're now able to estimate the ICA model, which means to uncover the hidden factors S and linear mixing matrix A in the equation (3.1). The key to estimating the ICA model is the nongaussianity assumption. In fact, without this assumption, the estimation is not possible as explained in the previous section.

The idea comes from the Central Limit Theorem (CLT). The CLT tells that the scaled sum of a large number of i.i.d. random variables tends to Gaussian in distribution, if the random variables have finite mean and variance. Thus, the scaled

sum of independent random variables has a distribution that is more Gaussian than the the original random variables.

Let the ICA model be defined as in Definition (3.1) and assume that the observed data X has the linear representation according to the model. Our goal is to find S and A. We multiply both sides of the equation by a matrix W,

$$Y = W'X = W'AS.$$

We further define $Z = A'W$, then, we have

$$Y = W'X = W'AS = Z'S.$$

Thus, Y is a linear sum of s_i. Since the sum of independent random variables is more Gaussian than the original variables, $Z'S$ is more Gaussian than any of the original s_i and is least Gaussian if it's equal to one of the original variable s_i. So, we want to find a matrix W such that we could maximizes the nongaussianity of $W'X$. Then, we have that

$$Y = W'X = W'AS \approx S,$$

which is implied by $W \approx A^{-1}$. Thus, we find the unknown source S and mixing matrix A.

3.5 Measurement of Nongaussianity

We know from the last section that the idea of estimating the ICA model is to maximize the nongaussianity of the linear transform of the observed data.

So, the question is how to measure the nongaussianity. One way to measure the nongaussianity is the kurtosis or the fourth order moment. But, it turns out to be an unrobust measure of nongaussianity [29].

Another candidate for the measure of nongaussianity is the negentropy. Negentropy is based on the quantity of entropy. Entropy is a basic concept of information theory. It describes how much randomness (or, alternatively, "uncertainty") there is in a random variable. The more "random" the variable is, the larger its entropy. A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance [17]. This fact tells that the Gaussian distribution is the "most random" or the least structured of all distributions. Thus, we can use entropy as a measure of the nongaussianity. Entropy H of a discrete random variable Y is defined as

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i),$$

where a_i are the possible values of Y. This definition can be generalized for continuous valued random variable. Under continuous case, it's called differential entropy, and is defined as

$$H(y) = -\int f(y) \log f(y) dy,$$

where $f(y)$ is the density function of the random variable y.

The negentropy J is a modified version of differential entropy and it's defined as follows

$$J(y) = H(y_{gauss}) - H(y),$$

where y_{gauss} is a Gaussian random variable, the variables y and y_{gauss} have the same covariance matrix.

Negentropy is always nonnegative by the fact that a Gaussian random variable has the largest entropy among all the random variables with the same variance. It is zero if and only if the random vector y is Gaussian. It's also invariant under invertible linear transformation. Finally, by introducing the concept of negentropy, we turn maximizing nongaussianity into maximizing negentropy. In fact, we use negentropy (differential entropy) as a measure of nongaussianity because it is in some sense the optimal estimator of nongaussianity [29].

However, it's not easy for us to use it directly. In practice, we would estimate it by approximations. In general, we have the following approximation:

$$J_g(y) \approx k \left[\mathbb{E}\{g(y)\} - \mathbb{E}\{g(y_{gauss})\}\right]^2, \tag{3.9}$$

where variable y is assumed to be of zero mean and unit variance, g is a nonquadratic function, k is some positive constant, and y_{gauss} is a standard Gaussian variable. Hyvärinen [26] gives the following practical choices of g:

$$g_1(x) = \frac{1}{\alpha_1} \log \cosh(\alpha_1 x), \tag{3.10}$$

$$g_2(x) = \frac{1}{\alpha_2} \exp(-\alpha_2 x^2/2), \tag{3.11}$$

with g' being their derivative:

$$g_1'(x) = \tanh(\alpha_1 x), \tag{3.12}$$

$$g_2'(x) = x \exp(-\alpha_2 x^2/2), \tag{3.13}$$

where α_1 and α_2 are constants. Experimentally, it was found that the choice of values $1 \leq \alpha_1 \leq 2$, and $\alpha_2 \approx 1$ gives good approximation.

3.6 Preprocessing the Data

Before we apply the algorithm of ICA to the data, it's very useful for us to do some preprocessing. This step will make estimating ICA model simpler and the original problem better conditioned.

3.6.1 Centering

The most basic preprocessing step is to center the observed data X, which means we subtract its mean vector $m = \mathbb{E}(X)$ from X:

$$X_{new} = X - \mathbb{E}(X) = AS - \mathbb{E}(X) = AS_{new}$$

After centering, both X_{new} and S_{new} are of zero mean, since

$$0 = \mathbb{E}(X - m) = \mathbb{E}(X_{new}) = A\mathbb{E}(S_{new}).$$

The mean of original unknown source $\mathbb{E}(S)$ is given by $A^{-1}m$ because

$$0 = \mathbb{E}(X) - m = A\mathbb{E}(S) - m \Longrightarrow A\mathbb{E}(S) = m.$$

From now on, we assume all the observed data have been centered. For notational simplicity, we still use the symbol X for the centered data.

3.6.2 Whitening

The next preprocessing step is to whiten the centered observed data. This step will linearly transform the data X into whitened data $\widetilde{X}$. By whitened data, we mean components of $\widetilde{X}$ are uncorrelated and their variances are equal to unity. Mathematically, the covariance matrix of $\widetilde{X}$ is an identity matrix:

$$\mathbb{E}(\widetilde{X}\widetilde{X}') = I.$$

One method for whitening the data is to apply the eigenvalue decomposition of covariance matrix C of $\widetilde{X}$:

$$\mathbb{E}(XX') = C = EDE', \tag{3.14}$$

where E is the orthogonal matrix of the eigenvectors of matrix C, D is the diagonal matrix of its eigenvalues, $D = diag(d_1, \cdots, d_n)$ and $D^{-\frac{1}{2}} = diag(d_1^{-\frac{1}{2}}, \cdots, d_n^{-\frac{1}{2}})$. We notice that if we take

$$\widetilde{X} = ED^{-\frac{1}{2}}E'X, \tag{3.15}$$

further, by applying Equation(3.14) and recalling $E'E = EE' = I$, we may check:

$$\begin{aligned}
\mathbb{E}(\widetilde{X}\widetilde{X}') &= \mathbb{E}\left((ED^{-\frac{1}{2}}E'X)(X'ED^{-\frac{1}{2}}E')\right) \\
&= (ED^{-\frac{1}{2}}E')\mathbb{E}(XX')(ED^{-\frac{1}{2}}E') \\
&= (ED^{-\frac{1}{2}}E')(EDE')(ED^{-\frac{1}{2}}E') \\
&= I.
\end{aligned}$$

By taking the linear transform, we have:

$$\widetilde{X} = ED^{-\frac{1}{2}}EX = ED^{-\frac{1}{2}}EAS = \widetilde{A}S, \tag{3.16}$$

where $\widetilde{A}$ is an orthogonal matrix. We may check this by recalling $\mathbb{E}(SS') = I$ and noting:

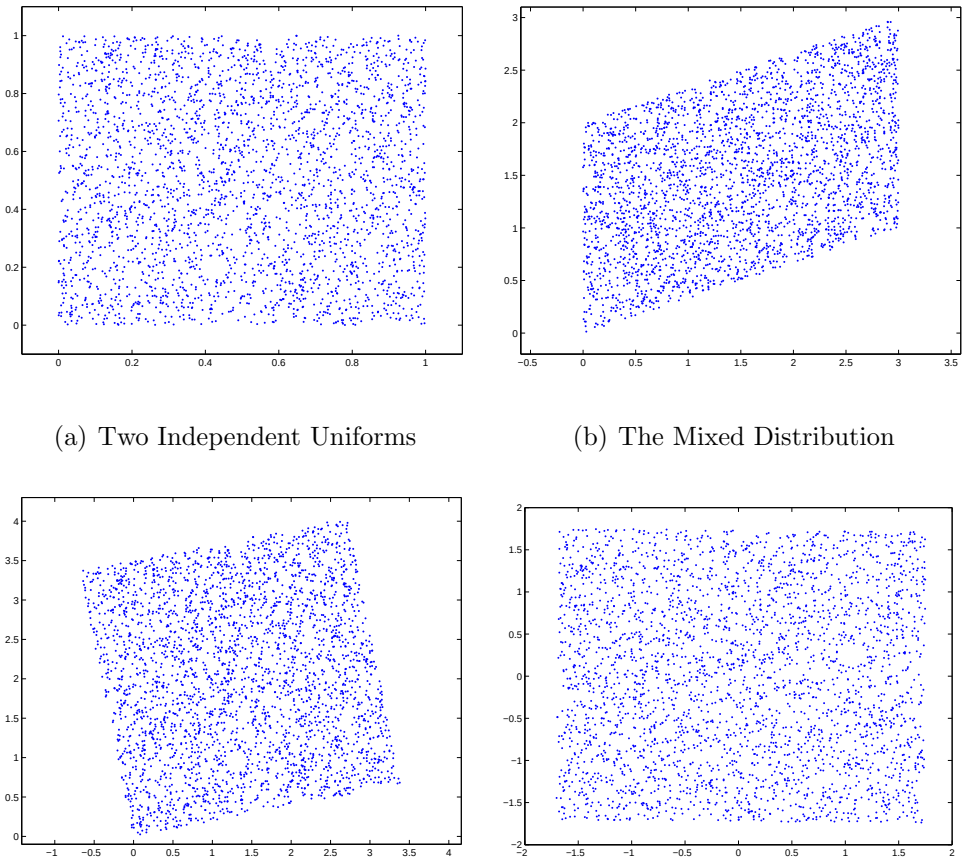
$$\mathbb{E}(\widetilde{X}\widetilde{X}') = \widetilde{A}\mathbb{E}(SS')\widetilde{A}' = \widetilde{A}\widetilde{A}' = I. \tag{3.17}$$

In fact, the purpose of whitening is to transform the original mixing matrix A into an orthogonal matrix $\widetilde{A}$. We know that orthogonal matrix only contains $n(n-1)/2$ degrees of freedom, which is less than n^2 parameters for the original matrix A. Thus, the whitening step reduces the number of unknown parameters and makes the ICA estimation simpler.

To illustrate the whitening effect and ICA, we present an illustration in Figure (3.1). We start with two independent components with uniform distribution, their joint distribution is shown as the upper-left picture. Then, we mix them with mixing matrix, their joint distribution is shown as the upper-right picture. The lower-left picture shows the joint distribution after whitening step. At last, we apply ICA to estimate the angle that gives the rotation, and this is shown in the lower-right picture.

3.7 The FastICA Algorithm

The FastICA algorithm is a fixed point iteration developed in [27]. More detailed treatment of FastICA algorithm may be found in [27, 29]. Here, we show the one unit version of the FastICA algorithm. For each iteration, the FastICA algorithm finds a vector w such that $w'x$ maximize nongaussianity measured by approximation of negentropy $J(w'x)$ defined in (3.9). The FastICA algorithm is



(a) Two Independent Uniforms

(b) The Mixed Distribution

(c) The Whitened Data

(d) Data Uncovered by ICA

Figure 3.1: Illustration for the effect of Whitening and ICA

1.	Center the data to make its mean zero,
2.	Whiten the centered data to get x,
3.	Choose randomly an initial weight vector w_0 of unit norm,
4.	Let $w^+ = E\{xg(w'x)\} - E\{g'(w'x)\}w$, where g and g' are defined in (3.10 - 3.13)
5.	Let $w_1 = w^+ / \parallel w^+ \parallel$,
6.	If not converged ($\|1 - w_0'w_1\| > \epsilon$), go back to 4.

Table 3.1: The FastICA Algorithm

given in Table (3.1).

We may derive the FastICA algorithm as follows. We start from an optimization problem:

$$\begin{aligned} &\text{Maximize} \qquad J_g(w'x) \\ &\text{such that} \qquad \mathbb{E}\{(w'x)^2\} = \| w \| = 1. \end{aligned} \tag{3.18}$$

By taking the gradient of $J_g(w'x)$ w.r.t w, we may translate the problem (3.18) into the following Lagrangian problem:

$$F(w) = \mathbb{E}\{xg(w'x)\} - \lambda w = 0, \tag{3.19}$$

where λ is the Lagrange multiplier. Now, we find the optima of Lagrangian by Newton's method. We get the Jacobian matrix $JF(w)$ as

$$JF(w) = \mathbb{E}\{xx'g'(w'x)\} - \lambda I. \tag{3.20}$$

To simplify the inversion of this matrix, we make the following approximation: $\mathbb{E}\{xx'g'(w'x)\} \approx \mathbb{E}\{xx'\}\mathbb{E}\{g'(w'x)\} = \mathbb{E}\{g'(w'x)\}I$. Then we obtain the Newton iteration:

$$w^+ = w - \left[\mathbb{E}\{xg(w'x)\} - \lambda w\right] / \left[\mathbb{E}\{g'(w'x)\} - \lambda\right]. \tag{3.21}$$

Finally, we multiply both sides of equation (3.21) by $\lambda - \mathbb{E}\{g'(w'x)\}$ to obtain the FastICA iteration.

Now, we give an example of estimating the ICA model. We start with two independent factors shown in the first row of the illustration. Then, we mix them and get the mixed observed data shown in the second row. The two pictures in third

row shows that the ICA uncover the original independent factors. The original data are generated by $sin(5t)$ and $cos(t)$, and the mixing matrix is given by:

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix}.$$

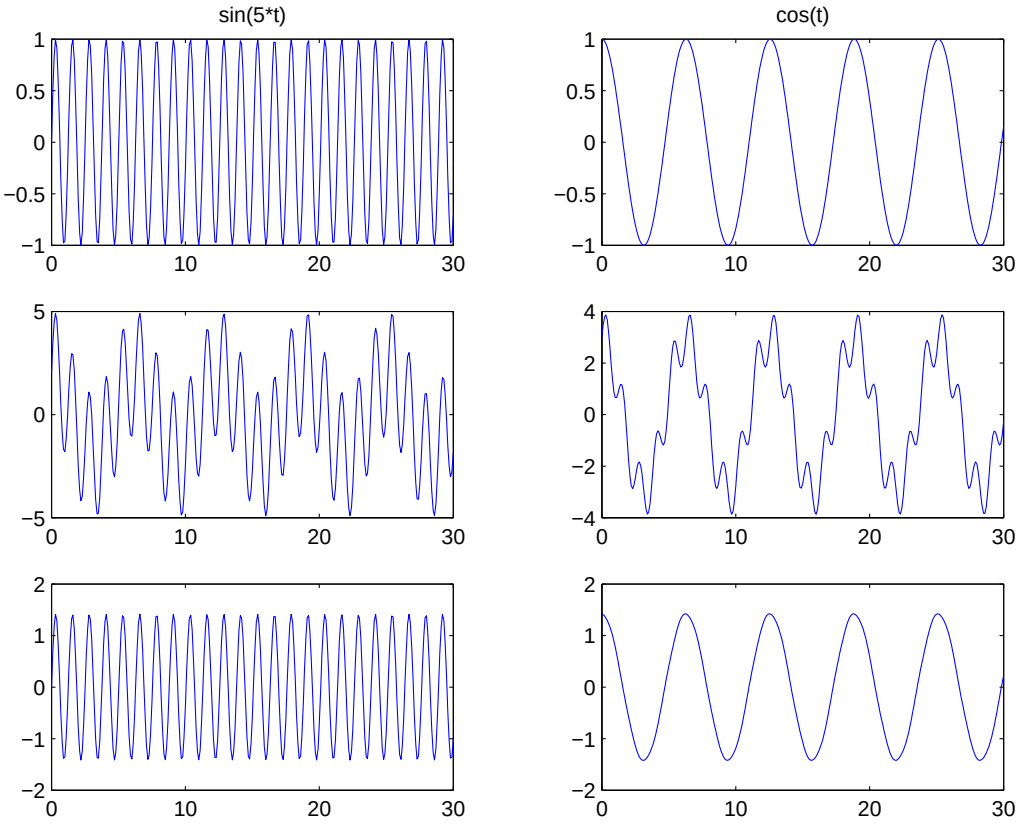


Figure 3.2: Illustration of Estimating ICA Model

Chapter 4

The VG Multiasset Option Pricing by ICA

The traditional Mean-Variance preference investment analysis was proposed by Markowitz [41] in 1952. It deals with individual agents on portfolio selection. The agents need to maximize the return measured by mean, whilst minimize the risk measured by the variance. The problem is modelled as a constrained optimization problem. With the Gaussian assumption and for some utility function, we may solve for the optimal portfolio.

The Gaussian assumption simplifies the analysis but also incurs problem. The presence of skewness preference and kurtosis aversion is observed in the analysis of single asset. But, we know that Gaussian models ignore the higher moments. To remedy this problem, we turn to the multidimensional counterpart of Lévy processes. However, multidimensional Lévy models are more difficult to construct than one-dimensional ones. The difficulties lie in the estimation of multidimensional probability laws and simulation of the multidimensional law.

Here we follow the perspective of independent components analysis(ICA) proposed by Madan and Yen [46]. We postulate the log-return of individual asset as a linear sum of some unknown factors driven by Lévy processes. Here, we treat the VG process as a special case of Lévy processes. We recover the independent factors and mixing matrix by applying ICA technique. We thus decompose the observed

time series data into statistically independent components (ICs) and translate the original multidimensional problem into multiple one-dimensional problems.

Once we recover the hidden factors, we calibrate their measures by Maximum Likelihood Estimation(MLE). However, the factors are estimated from the historical time series data, which means they follow the Physical law. To price option, we still need the risk-neutral law. We bridge the gap by introducing the measure change process, and estimate the measure change parameters through individual listed option data. Finally, we price multiasset products by doing Monte Carlo simulation.

4.1 Linear Representation of the Log-return

In this section, we explore extending the Variance Gamma model to multiasset product pricing. As we mentioned before, the extension is not direct due to the difficulty of handling multidimensional law. Here we follow Madan and Yen [46]'s framework, i.e., we postulate log-return as a linear mixture of independent factors driven by VG process.

To be more specific, suppose that we have a multiasset financial instrument consisting of n underlying assets $S_1, S_2, \cdots, S_n$. There're n independent factors driven by $F_1(t), F_2(t), \cdots, F_n(t)$, where $F_i(t)$ is a VG process. Then, each asset S_i is driven by the linear mixture of these factors:

$$\ln(S_i(t)/S_i(0)) = \mu_i t + \sum_{j=1}^{n} a_{ij} F_j(t), i = 1, 2, \cdots, n, \tag{4.1}$$

where $a_{ij}, 1 \leq i, j \leq n$, are weighting factors, and μ_i is the mean rate of return of

asset S_i under physical measure.

We may discretize the dynamics (4.1) and have the one period model. Let the log price increment (log-return) be X and the corresponding linear sum of increment of VG processes be Y, we'll have the linear representation form:

$$X = \mu + AY. \tag{4.2}$$

In the above representation form, the LHS of the equation contains the observable data, which are calculated from the historical time series data for each asset. The unknown items we want to uncover are the mixing matrix A and the hidden factors Y in the RHS of the equation. Here, we don't need to worry about the μ term, because it'll be replaced by interest rate less dividend rate in the risk-neutral pricing setting. Since the distribution of each unknown factor follows a nongaussian distribution, we can apply the ICA technique explained in Chapter 3 to recover A and $Y = (y_1, \cdots, y_n)'$.

Note in the ICA procedure, the preprocessing step will make $X - \mu$ in the equation (4.2), thus, Y has zero mean, its component y_i is a pure jump process and can be written as a Gamma time changed Brownian motion with parameters $(\theta_i, \sigma_i, \nu_i)$:

$$y_i = \theta_i(g_i - 1) + \sigma_i W_i(g_i), i = 1, 2, \cdots, n,$$

where g_i follows a Gamma distribution with unit mean and variance ν_i. In the following section, we'll discuss how to estimate the parameters for the VG distribution.

In our context, we assume that the unknown factors follow the VG law. Note that one could extend this framework to model the factors as any Lévy process

(except for Brownian motion!).

4.2 Maximum Likelihood Estimation of Parameters

The method of maximum likelihood(ML) was proposed by Fisher between 1912 and 1922 in three different papers. For the history of maximum likelihood, we refer to the paper by Aldrich [1]. Among all the point estimation methods, maximum likelihood estimation (MLE) is more preferred than any other methods, for example, the method of moments, the Bayesian estimator. Maximum likelihood estimator is an asymptotic efficient estimator for large sample. For our case, the factors are derived from historical asset prices data, which guarantees our data are of large sample size. So, we choose to work with MLE to estimate the parameters for the data.

4.2.1 Method of Maximum Likelihood

Suppose we have an i.i.d. sample $x_i, i = 1, \cdots, n$ from a population with probability density function $f(x_i|\boldsymbol{\theta})$, conditioned on a set of parameters $\boldsymbol{\theta}$. The likelihood function of the sample is given by

$$l(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

The MLE will find a set of parameters to maximize the likelihood function. Usually, we work with minimizing the negative of the log of the likelihood function

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \ln(l(\boldsymbol{\theta}|\boldsymbol{x})).$$

One important property of maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is the asymptotic normality and efficiency. It tells that as the sample size increases, the sampling distribution of $\hat{\boldsymbol{\theta}}$ converges to normal:

$$\hat{\boldsymbol{\theta}} \rightarrow N[\boldsymbol{\theta}_0, I(\boldsymbol{\theta}_0)^{-1}],$$

where $\boldsymbol{\theta}_0$ is the true value of the parameters, and $I(\boldsymbol{\theta})$ is the Fisher information matrix defined as:

$$I(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right].$$

We notice that the variance of $\hat{\boldsymbol{\theta}}$ is given by the inverse of the Fisher information matrix. In fact, $\hat{\boldsymbol{\theta}}$ achieves the Cramér-Rao lower bound for consistent estimator, which means it generally has smaller variance than other estimator.

For large sample sizes, Fisher information matrix is often approximated by the Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}$:

$$I(\boldsymbol{\theta}) \approx -\boldsymbol{H}(\hat{\boldsymbol{\theta}}),$$

where

$$\boldsymbol{H}_{ij}(\boldsymbol{\theta}) = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}.$$

4.2.2 MLE for the VG Processes

As we see from last section, we need to know the pdf function of the sample data to apply the method of maximum likelihood. For VG process at time t, we notice that its conditional distribution (on Gamma time change g) has normal density. The unconditional density function is then obtained by integrating out g [42].

We have the VG distribution at time t expressed as follow:

$$f_t(x) = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}} \exp\left(-\frac{(x-\theta g)^2}{2\sigma^2 g}\right) \frac{g^{\frac{t}{\nu}-1}\exp\left(-\frac{g}{\nu}\right)}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})} dg. \tag{4.3}$$

The pdf for VG is expressed as an integration and it's difficult for us to find the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ analytically. In practise, we need to numerically find its density function to apply the method of maximum likelihood. We further notice that its characteristic function for unit time is given by :

$$\mathbb{E}[\exp(iuX)] = \int_{-\infty}^{\infty} \exp(iux) f(x) dx = \phi(u),$$

So, we may calculate the pdf by inverting $\phi(u)$ through inverse Fourier transform

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-iux)\phi(u) du.$$

For the characteristic function $\phi(u)$, we recall that

$$x = \theta(g-1) + \sigma W(g), \tag{4.4}$$

where W is a Brownian motion, and g follows a Gamma distribution with unit mean and variance ν. Then, the characteristic function of x is

$$\phi(u) = \mathbb{E}(e^{iux}) = \exp\left(-\frac{1}{\nu}\ln\left(1 - iu\theta\nu + \frac{\sigma^2\nu}{2}u^2\right) - iu\theta\right). \tag{4.5}$$

4.3 The Measure Change for Pure Jump Processes

The change of measure plays an important role in our proposed method to price multiasset option driven by VG. Recall that our factors are derived from log-returns for each asset. The data are coming from the historical stock prices. Thus,

the measures of factors identified by MLE are actually physical measures. We know that to price option, we need to apply the risk-neutral measure. In this section, we bridge the gap between physical measure and risk-neutral measure by introducing the change of measure process for Lévy process. And we'll focus on the change of measure for pure jump processes.

In the case of measure change for pure jump processes, we'll change the distribution of large jumps. Jacod and Shiryaev [34] show that the change of measure process for a pure jump process can be explicitly computed from its physical ($\mathbb{P}$) and risk-neutral ($\mathbb{Q}$) Lévy measures. Suppose that we have the pure jump processes with Lévy density $k_P(x)$ and $k_Q(x)$ under the $\mathbb{P}$ and $\mathbb{Q}$ measures, respectively. If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures, the Radon-Nikodym derivative is given by

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \exp\left(-t\int_{-\infty}^{\infty}(Z(x)-1)k_P(x)dx\right)\prod_{s\leq t}Z(\Delta X(s)), \tag{4.6}$$

where $Z(x)$ is given by

$$k_Q(x) = Z(x)k_P(x).$$

The previous formula 4.6 reveals the relationship between physical measure and risk-neutral measure expressed by the measure change process. Given the explicit form of measure change, we may further infer the measure change from the information we know about both measures.

4.4 Tilting the Lévy Measure

In practice, the transformation of Lévy measures are usually given by multiplying the Lévy measure $\nu(dx)$ by an exponential function

$$\tilde{\nu}(dx) = e^{\theta x}\nu(dx), \tag{4.7}$$

where $\theta \in \mathbb{R}$, and $\tilde{\nu}(dx)$ is the new Lévy measure. The exponential function is chosen such that

$$\int_{|x|\geq 1} e^{\theta x}\nu(dx) < \infty.$$

The transform (4.7) is called exponential tilting of the Lévy measure. The new Lévy process $\tilde{X}_t, t \geq 0$ with generating triplet $(\gamma, \sigma, \tilde{\nu})$ is also a Lévy process, and is called the Esscher transform of $X_t, t \geq 0$. Here, $X_t, t \geq 0$ is the original Lévy process with generating triplet (γ, σ, ν).

The exponential function in the tilting formula (4.7) infers the risk premia for jumps. Carr *et al.* [9] observe the different risk premium levels for positive and negative jumps in log price relative. To be able to capture the asymmetric U shaped measure change, we adopt the two sided tilting as well.

Back to our problem, we have the Lévy density $k_P(x)$ under $\mathbb{P}$ measure and $k_Q(x)$ under $\mathbb{Q}$ measure. We explicitly take the measure change function $Y(x)$ as follows:

$$Y(x) = \begin{cases} \gamma \exp(\alpha x - \beta|x|), & x < -\epsilon \\ \\ \gamma \exp(\alpha x - \beta|x|), & x > \epsilon \end{cases} \tag{4.8}$$

Note that we have the ϵ in the equation 4.8. Here, we assume that $Y(x) = 1, -\epsilon <$

$x < \epsilon$, which means we approximate $K_Q(x)$ by $K_P(x)$ for an arbitrary small interval including 0. Then, the risk-neutral Lévy density for unknown factor is given by

$$k_Q(x) = Y(x)k_P(x). \tag{4.9}$$

We next need to calibrate the measure change parameters α, β, γ from physical and risk-neutral measure.

4.5 Risk-neutral Stock Price Dynamics

In this section, we'll derive the risk-neutral stock price dynamics driven by n VG factors. Let's first take a look at the one factor case. Suppose the risk-neutral stock price was driven by one VG process scaled by a constant $a \in \mathbb{R}$. Then, the risk-neutral VG stock price dynamics is obtained by the exponential function of the drift term, the scaled VG process, and the a convexity correction factor to make the discounted stock price a martingale.

Proposition 4.1 *Let $S(t)$ be the asset price dynamics driven by a scaled VG process $Y(t;\sigma,\nu,\theta)$, then, the risk-neutral VG stock price dynamics is given by*

$$S(t) = S(0)\exp((r-q)t + aY(t;\sigma,\nu,\theta) + \omega t), \tag{4.10}$$

where r is the continuously compounded interest rate, q is the dividend yield rate, if the law of $Y(t;\sigma,\nu,\theta)$ is

$$Y(t;\sigma,\nu,\theta) = \theta(g-t) + \sigma W(g), \tag{4.11}$$

where g follows Gamma distribution with mean t and variance νt. Then, the con-

vexity correction factor

$$\omega = \phi_{VG}(a/i) + a\theta = \frac{1}{\nu}\ln(1 - a\theta\nu - \frac{1}{2}a^2\sigma^2\nu) + a\theta. \tag{4.12}$$

Recall our linear representation form, the log-return for each asset is the linear sum of n factors driven by VG processes with risk-neutral parameters. Thus, in terms of n independent factors, we apply Proposition 4.1 and we have risk-neutral VG stock price dynamics given as the following theorem.

Theorem 4.2 *Let $S_i(t)$ be the asset price dynamics driven by n scaled VG processes $Y_i(t;\sigma_i,\nu_i,\theta_i)$, then, the risk-neutral VG stock price dynamics is given by*

$$S_i(t) = S_i(0)\exp\left((r - q_i)t + \sum_{j=1}^{n} a_{ij}Y_j(t) + \sum_{j=1}^{n}\omega_{ij}t\right), \tag{4.13}$$

where r is the continuously compounded interest rate, q_i is the dividend yield rate for asset i, if the law of $Y(t;\sigma_j,\nu_j,\theta_j)$ is

$$Y(t;\sigma_i,\nu_j,\theta_j) = \theta_j(g - t) + \sigma_j W(g), \tag{4.14}$$

where g follows Gamma distribution with mean t and variance νt. Then, the convexity correction factor

$$\omega_{ij} = \frac{1}{\nu_j}\ln(1 - a_{ij}\theta_j\nu_j - \frac{1}{2}a_{ij}^2\sigma_j^2\nu_j) + a_{ij}\theta_j. \tag{4.15}$$

Next, we derive the characteristic function of the log-return for each asset. Before that, we'll first introduce two Propositions.

Proposition 4.3 *Let $Y(t;\sigma,\nu,\theta)$ be a VG process with zero mean, $a \in \mathbb{R}$, and g be a Gamma random variable with mean t and variance νt. Then, the scaled VG*

process $aY(t;\sigma,\nu,\theta)$ *has the characteristic function:*

$$\begin{aligned}\mathbb{E}(\exp(iuaY_t)) &= \mathbb{E}\left(\exp(iau(\theta g+\sigma W(g)-\theta t)\right) &(4.16)\\ &= \phi_t(au)\exp(-iau\theta t), &(4.17)\end{aligned}$$

where $\phi_t(u)$ *is the characteristic function of the VG process (see formula 2.13).*

Proposition 4.4 *Let* $Z_1, Z_2, \cdots, Z_n$ *be independent random variables, and* c_1, c_2,$\cdots$, $c_n \in \mathbb{R}$ *are constant, then the characteristic function of linear sum of them will be the product of their characteristic function* $\phi_Z(u)$*. Mathematically,*

$$\mathbb{E}\left(\exp\left(iu\sum_i^n c_jZ_j\right)\right)=\prod_i^n \phi_{Z_j}(c_ju). \tag{4.18}$$

Now, we derive the characteristic function for the log-return. By our assumption, the log-return is the linear mixture of independent VG factors. Apply Theorem 4.2, Proposition 4.3 and 4.4, we'll have the following theorem.

Theorem 4.5 *Let* x_i *be the log of* $S_i(t)$*, and* $r, q_i, Y_j, \omega_{i,j}$ *are defined as before. Then, the characteristic function of log price at time* t *is given by*

$$\phi_{x_i}(u,t)=\exp\left[iu(\ln(S_i(0))+(r-q_i+\tilde{\omega}_i)t)\right]\prod_{j=1}^n \phi_{Y_j(t)}(a_{ij}u), \tag{4.19}$$

where

$$\tilde{\omega}_i=\sum_j^n \frac{1}{\nu_j}\ln(1-a_{ij}\theta_j\nu_j-\frac{1}{2}a_{ij}^2\sigma_j^2\nu_j).$$

proof We first rewrite equation (4.13) in terms of log-return

$$x_i=\ln(S_i(t))=\ln(S_i(0))+(r-q_i)t+\sum_{j=1}^n a_{ij}Y_j(t)+\sum_{j=1}^n \omega_{ij}t.$$

then, we calculate the characteristic functions of both sides, and by apply Theorem 4.2, Proposition 4.3 and 4.4:

$$\begin{aligned}
\phi_{x_i}(u,t) &= \mathbb{E}(e^{iux_i}) \\
&= \mathbb{E}\left[\exp\left(iu\left(\ln(S_i(0)) + (r-q_i)t + \sum_{j=1}^{n} a_{ij}Y_j(t) + \sum_{j=1}^{n}\omega_{ij}t\right)\right)\right] \\
&= \exp\left[iu(\ln(S_i(0)) + r - q_i)t\right]\prod_{j=1}^{n}\phi_{Y_j(t)}(a_{ij}u)\exp\left(\sum_{j=1}^{n} iu\left(-a_{ij}\theta_j + \omega_{ij}\right)t\right) \\
&= \exp[iu(\ln(S_i(0)) + (r-q_i)t)]\prod_{j=1}^{n}\phi_{Y_j(t)}(a_{ij}u)\exp(iu\tilde{\omega}_i t) \\
&= \exp[iu(\ln(S_i(0)) + (r-q_i+\tilde{\omega}_i)t)]\prod_{j=1}^{n}\phi_{Y_j(t)}(a_{ij}u)
\end{aligned}$$

where

$$\tilde{\omega}_i = \sum_{j}^{n}\frac{1}{\nu_j}\ln(1 - a_{ij}\theta_j\nu_j - \frac{1}{2}a_{ij}^2\sigma_j^2\nu_j).$$

Thus, complete the proof. ■

Then, the call option price for asset i with maturity T can be expressed as:

$$C_T(k) = \frac{\exp(-\alpha k)}{\pi}\int_0^{\infty} e^{-ivk}\psi_T(v)dv, \tag{4.20}$$

where $\exp(-\alpha k)$ is the dampening factor, $\psi_T(v)$ is defined as follows:

$$\psi_T(v) = \frac{e^{-rT}\phi_{x_i}\left(v-(\alpha+1)i, T\right)}{\alpha^2 + \alpha - v^2 + iv(2\alpha+1)}. \tag{4.21}$$

Finally, we may calibrate the measure change parameters simultaneously through individual listed option prices. With the measure change parameters for a fixed maturity T, we may price the multiasset option across different strikes.

4.6 Numerical Implementation

We are considering a portfolio with 10 stocks. By ticker symbol, the 10 stocks are: BA, C, GE, IBM, JNJ, KO, MRK, ORCL, PFE, and WMT. We postulate that log-return of each asset is the linear mixture of Independent Components following Variance Gamma process, as expressed in equation (4.2). The time series data are taken from Jan. 2nd, 1998 to Apr. 23rd, 2003. We calibrated the VG parameters by MLE and the results are presented in the following Table 4.1.

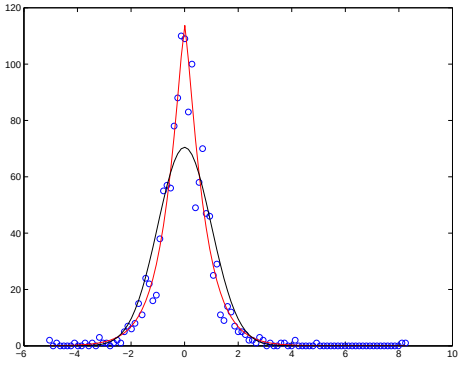
ICs	σ	ν	θ
1	0.9375	0.3407	-0.0096
2	0.9372	0.6944	-0.0040
3	0.9155	0.4074	-0.3579
4	0.9647	0.4874	0.1686
5	0.9531	0.6326	0.0006
6	0.9967	0.6018	-0.1024
7	0.9941	0.5324	-0.0273
8	0.9928	0.3932	-0.0013
9	0.9921	0.4508	0.1165
10	0.9912	0.2865	-0.2186

Table 4.1: Statistical Measure

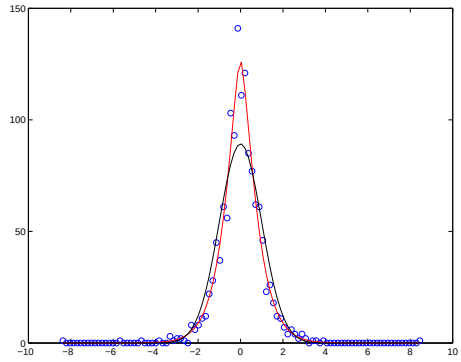
We also list the estimated density fitting for 10 independent components of the 10 asset, where the circles denote the binned data, the red line denotes the VG fitting and the black line denotes the Gaussian fitting.

We next present the results for calibration of measure change parameters. Recall that for each unknown factor, we have

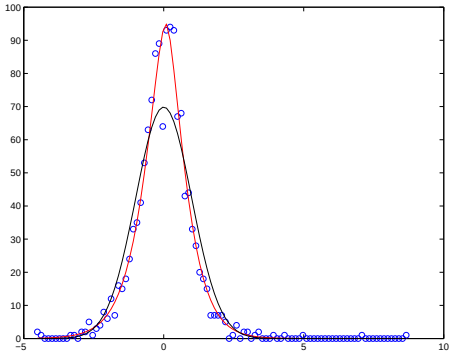
$$k_Q^i(x) = Y_i(x)k_P^i(x), i = 1, 2 \dots n, \tag{4.22}$$



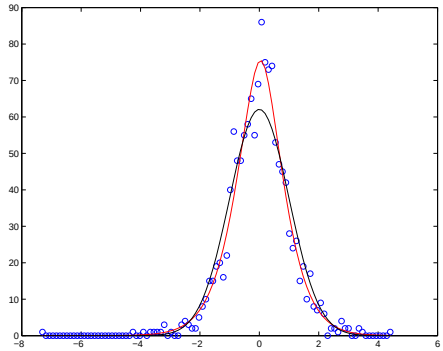
(a) The 1st IC Density Fitting



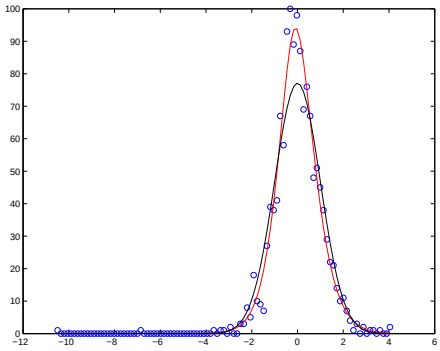
(b) The 2nd IC Density Fitting



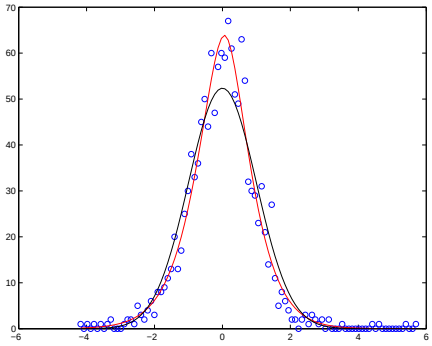
(c) The 3rd IC Density Fitting



(d) The 4th IC Density Fitting

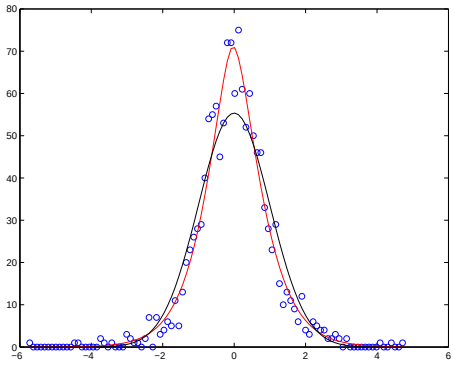


(e) The 5th IC Density Fitting

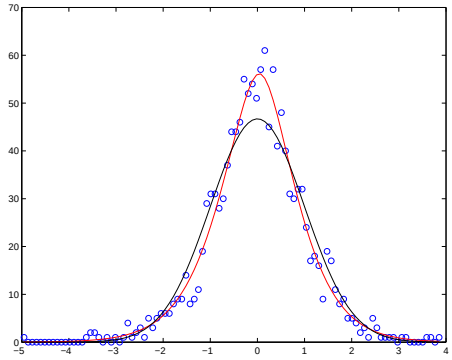


(f) The 6th IC Density Fitting

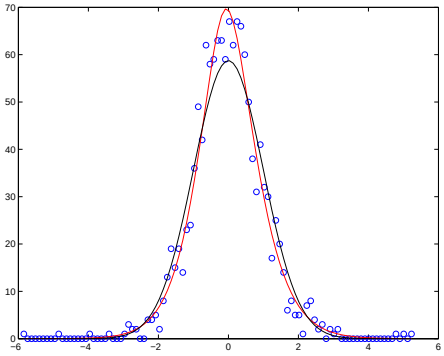
Figure 4.1: Independent Components Density Fit



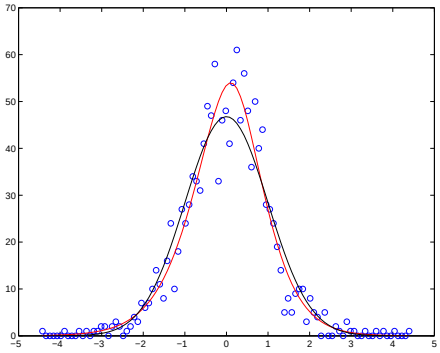
(a) The 7th IC Density Fitting



(b) The 8th IC Density Fitting



(c) The 9th IC Density Fitting



(d) The 10th IC Density Fitting

Figure 4.2: Independent Components Density Fit (con't)

where

$$Y_i(x) = \begin{cases} \gamma_i \exp(\alpha_i x - \beta_i |x|), & x < \epsilon \\ \gamma_i \exp(\alpha_i x - \beta_i |x|), & x > \epsilon \end{cases} \tag{4.23}$$

Based upon our proposed explicit form of measure change, we could calibrate the measure change parameters simultaneously through the listed option data. Option data are taken on Jan. 15th, 2003 and we looked at 3 maturities 38, 367, 738 days. As of the measure of fit, the Absolute Percentage Errors (APE) are used

$$APE = \frac{\sum_{i=1}^{N} |P_i - \tilde{P}_i|}{\sum_{i=1}^{N} P_i}, \tag{4.24}$$

where $P_i, \tilde{P}_i, i = 1, \cdots, N$ are the model prices and market prices respectively. Here, we list the fitting errors for different maturities in Table 4.6.

Maturity	38 Days	367 Days	738 Days
APE	0.0101	0.0159	0.0235
Number of Data	107	109	93

Table 4.2: Fitting Errors

The calibrated Measure Change Parameters for 38 days, 367 days and 738 days are listed in Table 4.3, 4.4 and 4.5.

The measure change density for equity has the asymmetric U shape [9]. In our case, we observe the following:

$$\begin{cases} \alpha_i - \beta_i > 0, & x > 0 \\ \alpha_i + \beta_i < 0, & x < 0 \end{cases}, i = 1, 2...10, \tag{4.25}$$

which means that the measure change density also keeps the U shape. We may explain the fact as follows: Recall that the log-return is postulated as the linear

ICs	α	β	γ
1	0.2156	-3.0233	4.7240
2	0.3391	-1.6966	6.5510
3	0.2475	-2.9028	5.1931
4	0.3827	-1.8439	2.4744
5	1.0212	-2.7512	2.0191
6	0.7939	-3.3624	1.9291
7	2.0836	-1.9702	2.0567
8	0.4787	-2.0812	4.5828
9	0.6802	-2.7193	2.5105
10	0.7214	-2.5155	6.6464

Table 4.3: Measure Change Parameters (38 Days)

ICs	α	β	γ
1	-0.0596	-2.4628	0.2917
2	-0.3367	-2.2088	0.5952
3	0.5911	-2.1460	2.1289
4	0.8237	-2.9836	1.4214
5	0.1350	-2.1774	2.2861
6	0.3360	-3.9075	2.4441
7	0.0967	-3.9057	2.5799
8	0.6933	-1.3985	2.9537
9	-1.5705	-3.5047	0.2626
10	0.5419	-2.2832	0.5713

Table 4.4: Measure Change Parameters (367 Days)

ICs	α	β	γ
1	-0.2909	-2.2432	0.2252
2	-0.8190	-2.6704	0.2865
3	0.7511	-2.0364	1.0990
4	0.3887	-2.6300	0.2911
5	0.2492	-2.2474	1.4480
6	0.0461	-3.6637	0.7405
7	0.5791	-1.8405	1.2931
8	0.3197	-1.8783	0.5545
9	0.5170	-1.5457	0.1390
10	0.6565	-2.1226	1.1798

Table 4.5: Measure Change Parameters (738 Days)

mixture of independent components. The U shaped measure change between physical measure and risk-neutral measure could be decomposed into U shaped measure change for individual components. Now, with these measure change parameters, we are able to price the multiasset options by the risk-neutral measure inferred by the measure change process. And we will present an example in the next chapter.

Chapter 5

The VG Multiasset Option Pricing by Copula Method

For a long time, statisticians have been interested in the problem of the relationship between the multivariate distribution function and the fixed univariate marginal distributions. Sklar [55] solved this problem by showing that all the finite dimensional probability laws have a copula function associated with them. The copula function virtually introduces the dependence structure for arbitrary marginal distributions to form a multivariate distribution. We refer to [47, 20, 15, 31] for detailed introduction to copula method.

Before the emerging of alternate copula based models, the study of multivariate time series has been dominated by elliptic models, like multivariate normal or t distributions. Their popularity only results from their mathematical tractability and was questioned by empirical financial data. For example, they can't model the co-occurrence extreme events and they assume the linear association between different assets. The alternating copula method could take any marginal distribution, which of course, accommodates non-gaussian distributions. Furthermore, the copula based models could assume non-linear association and tail dependence.

The copula based models was first introduced to model default correlation in credit field by Li [38] in 2000. Later, we see applications of copula methods in risk management [21, 20], in option pricing by [13, 25], in credit derivative pricing

like portfolio credit default swap (CDS) and Collateralized Debt Obligations (CDO) [5, 43, 4, 40].

5.1 Definitions of Copulas

The word copula is a Latin noun for link, something that ties things together. The name was chosen by the fact that a copula "couples" the marginal distribution functions to form multivariate distribution functions. Sklar [55] first used this word in his paper in 1959. The study of copulas was originally involved in the development of the theory of probabilistic metric spaces [54]. Later, more attentions are paid to the study of dependence structure and the construction of families of multivariate distribution [47].

The simple definition of a copula will be a multivariate distribution function defined on the n-dimensional unit hypercube $[0,1]^n$, with uniformly distributed marginal distribution functions. This definition is intuitive but not mathematically rigid, since it does not say what is meant by a multivariate distribution function and how to construct it. We'll give a mathematically rigid definition below. To begin with, we first define the notions of grounded function and n-increasing property, which will make copulas have the properties shared by distribution functions.

Throughout this chapter, we denote nonempty subsets of $\mathbb{R}$ by $S_1, S_2, \cdots, S_n$, we denote by $DomH$ and $RanH$ the domain and range respectively of a function H.

Definition 5.1 *Let H be a real function defined on $DomH = S_1 \times S_2 \times \cdots \times S_n$,*

$H : DomH \to \mathbb{R}$. Let a_i be the least element of $S_i, i = 1, 2, \cdots, n$. The function H is grounded if for all the elements $t_i \in S_i, i = 1, 2, \cdots, n$, we have:

$$H(a_1, t_2, \cdots, t_n) = H(t_1, a_2, t_3, \cdots, t_n) = \cdots = H(t_1, \cdots, t_{n-1}, a_n) = 0. \tag{5.1}$$

Definition 5.2 *Let H be a real function defined on $DomH = S_1 \times S_2 \times \cdots \times S_n$. Let $a_i, b_i \in S_i, a_i < b_i, i = 1, 2, \cdots, n$ and $B(= [a_1, b_1] \times \cdots \times [a_n, b_n])$ be an n-box. Then, the H-volume of B is given by:*

$$V_H(B) = \prod_{i=1}^{n} H_{a_i}^{b_i}(t), \tag{5.2}$$

where

$$H_{a_i}^{b_i}(t) = H(t_1, \cdots, t_{i-1}, b_i, t_{i+1}, \cdots, t_n) - H(t_1, \cdots, t_{i-1}, a_i, t_{i+1}, \cdots, t_n).$$

Definition 5.3 *Let H be a real function defined on $DomH = S_1 \times S_2 \times \cdots \times S_n$. Let $a_i, b_i \in S_i, a_i < b_i, i = 1, 2, \cdots, n$. A real function H of n variables is n-increasing if $V_H(B) > 0$ for all n-boxes B whose vertices lie in $DomH$.*

Proposition 5.4 *If H is a grounded n-increasing function, then H is increasing in each argument, i.e., if $(t_1, \cdots, t_{i-1}, x, t_{i+1} \cdots, t_n)$ and $(t_1, \cdots, t_{i-1}, y, t_{i+1} \cdots, t_n) \in DomH$, and $x \leq y$, then,*

$$H(t_1, \cdots, t_{i-1}, x, t_{i+1} \cdots, t_n) \leq (t_1, \cdots, t_{i-1}, y, t_{i+1} \cdots, t_n).$$

We refer to [54] for the proof. And we next formally define the copula.

Definition 5.5 *An n-dimensional copula is a function $C : [0, 1]^n \to [0, 1]$, which has the following properties :*

1. *C is grounded and n-increasing.*

2. $C(1, \cdots, 1, u_i, 1, \cdots, 1) = u_i, i = 1, \cdots, n, u_i \in [0, 1]$.

This definition ensures that the copula C can be used as a distribution function on $[0, 1]^n$. We also notice that copula C is a multivariate distribution function with uniformly distributed margins:

$$C(u_1, \cdots, u_n) = \mathbb{P}(U_1 \leq u_1, \cdots, U_n \leq u_n).$$

This also tells that copula function only work for random variables, and it's not for processes. And we have the following theorem:

Theorem 5.6 *Let C be an n-copula. Then for $\boldsymbol{u}$ and $\boldsymbol{v} \in [0, 1]^n$,*

$$|C(\boldsymbol{u}) - C(\boldsymbol{v})| \leq \sum_{i=1}^{n} |u_i - v_i|. \tag{5.3}$$

Hence, C is uniformly continuous on $[0, 1]^n$.

We refer to [54] for the proof. Another important result is the Fréchet-Hoeffding bounds. Consider multivariate distribution functions C^+, C^- defined on $[0, 1]^n$ as follows:

$$\begin{aligned} C^+(\boldsymbol{u}) &= \min(u_1, u_2, \cdots, u_n), \\ C^-(\boldsymbol{u}) &= \max\left(\sum_{i=1}^{n} u_i - n + 1, 0\right). \end{aligned}$$

Then, n-copula C is bounded:

$$C^-(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq C^+(\boldsymbol{u}),$$

we call $C^-(\boldsymbol{u})$ the Fréchet-Hoeffding lower bound and call $C^+(\boldsymbol{u})$ the Fréchet-Hoeffding upper bound.

Definition 5.7 *A pair of random variables X and Y are said to be comonotonic if $C(X,Y) = C^+$; if $C(X,Y) = C^-$, they are said to be countermonotonic.*

Here, comonotonic corresponds to perfectly positive dependent, and countermonotonic corresponds to perfectly negative dependent.

The following Theorem tells that the dependence structure described by copula is invariant under increasing and continuous transformations of the marginals.

Theorem 5.8 *If $(X_1, \cdots, X_n)$ has copula C and $T_1, \cdots, T_n$ are increasing continuous functions, then $(T_1(X_1), \cdots, T_n(X_n))$ also has copula C.*

We refer to [21] for the proof. This invariant property for copula is an attractive feature not shared by other dependence measure like linear correlation, explained in following section.

5.2 Copulas and Multivariate Distribution

In this section, we'll introduce an important theorem, which is the foundation of the application of copula in probabilistic modelling. The theorem tells that the construction of any multivariate distribution can be split into the construction of the marginals and the copula separately.

Theorem 5.9 (Sklar) *Let H be an n-dimensional multivariate distribution function with margins $F_1, F_2, ..., F_n$. Then there exists an n-dimensional copula C such that, for $x \in R^n$ we have*

$$H(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \tag{5.4}$$

If F_i are continuous, then C is unique; otherwise, C is uniquely determined on $\text{Ran}F_1 \times \text{Ran}F_2 \times \cdots \times \text{Ran}F_n$. *Conversely, if C is an n-copula and $F_1, \cdots, F_n$ are margins, then the function H defined in equation (5.4) is an n-dimensional multivariate distribution function with margins $F_1, \cdots, F_n$.*

The Sklar's theorem states that for continuous multivariate distribution functions, the multivariate dependence structure and margins can be separated from the multivariate distribution, and we can fully describe the dependence structure by copula function. Thus, given the marginal distribution functions and the dependence structure described by copulas, we may form the multivariate distribution function.

Definition 5.10 *Let F be a univariate distribution function. We define the generalized inverse of F as follow:*

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t, \text{for all } t \in [0,1]\} \tag{5.5}$$

Corollary 5.11 *Let H be an n-dimensional multivariate distribution function with continuous margins $F_1, \cdots, F_n$ and n-copula C. Then for any $\boldsymbol{u} \in [0,1]^n$,*

$$C(u_1, u_2, \cdots, u_n) = H(F_1^{-1}(u_1), F_2^{-1}(u_2), \cdots, F_n^{-1}(u_n)).$$

We refer to [56] for the proofs. This Corollary gives the Sklar's Theorem in the equivalent form.

In summary, a copula is a function such that, when applied to marginals, we form a proper multivariate distribution function. The resulting multivariate function contains the information about the dependence structure of its components. Thus, we break down the multivariate distribution of a random vector to individual components (marginals) with a dependence structure (the copula) between them.

5.3 Measures of Dependence

To describe the multivariate distribution, we need the measures of the dependence between random variables. Linear correlation (Pearson's correlation) is commonly used in practice to meet such needs. It works fine for elliptical distributions, but there's problems when it comes to non-elliptical margins. Thus, we'll discuss a copula-based measure of dependence: concordance measures and tail dependence. They are used to describe the global trend (concordance) and asymptotical dependence between extreme events (tail dependence).

5.3.1 Linear Correlation

We start with pairs of real-valued, non-degenerate random variables X,Y with nonzero finite variances.

Definition 5.12 *The linear correlation coefficient between X and Y is*

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \tag{5.6}$$

where $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ is the covariance between X and Y, and $Var(X)$ and $Var(Y)$ are the variances of X and Y.

The Linear correlation is a measure of linear dependence. Let $a, c \in \mathbb{R}\backslash\{0\}, b, d \in \mathbb{R}$, we notice the following linearity property:

$$\rho(aX+b, cY+d) = \text{sgn}(ac)\rho(X,Y) \tag{5.7}$$

Applying the above property, one may show that linear correlation is invariant under positive affine transformation (increasing linear transformation). We may

also extend equation (5.7) to higher dimension cases. Let A, C be $n \times n$ matrices, $a, b \in \mathbb{R}^n$ and let $\boldsymbol{X}, \boldsymbol{Y}$ be random n-vectors, we have:

$$Cov(A\boldsymbol{X} + b, C\boldsymbol{Y} + d) = ACov(\boldsymbol{X}, \boldsymbol{Y})C'. \tag{5.8}$$

The popularity of linear correlation comes from the fact that it's easy to calculate and manipulate under affine transformations. It works perfectly for multivariate normal distribution. But, when it comes to margins other than normal, there're problems with linear correlation. One problem comes up for margins with undefined variances like Cauchy distribution and student-t_2 distribution. Another problem is that linear correlation is not invariant under non-linear increasing transforms. Let such a transformation $T : \mathbb{R} \to \mathbb{R}$. For two real-valued random variables X and Y, we have

$$\rho(T(X), T(Y)) \neq \rho(X, Y). \tag{5.9}$$

Consider the following example [57]:

Proposition 5.13 *Let X and Y be two normal random variables with parameters (μ_X, σ_X),(μ_Y, σ_Y). Let ρ be the correlation between X and Y. Let $\widetilde{X} = \exp(X)$ and $\widetilde{Y} = \exp(Y)$. Then, the correlation between them:*

$$\rho(\widetilde{X}, \widetilde{Y}) = \frac{\exp(\rho\sigma_X\sigma_Y) - 1}{\sqrt{(\exp(\sigma_X^2) - 1)(\exp(\sigma_Y^2) - 1)}}. \tag{5.10}$$

See Appendix for the proof.

We fix the correlation between X and Y to be 1, and the upper and lower bounds for the correlation are:

$$\rho_{\max} = \frac{\exp(\sigma_X\sigma_Y) - 1}{\sqrt{(\exp(\sigma_X^2) - 1)(\exp(\sigma_Y^2) - 1)}} \geq 0$$

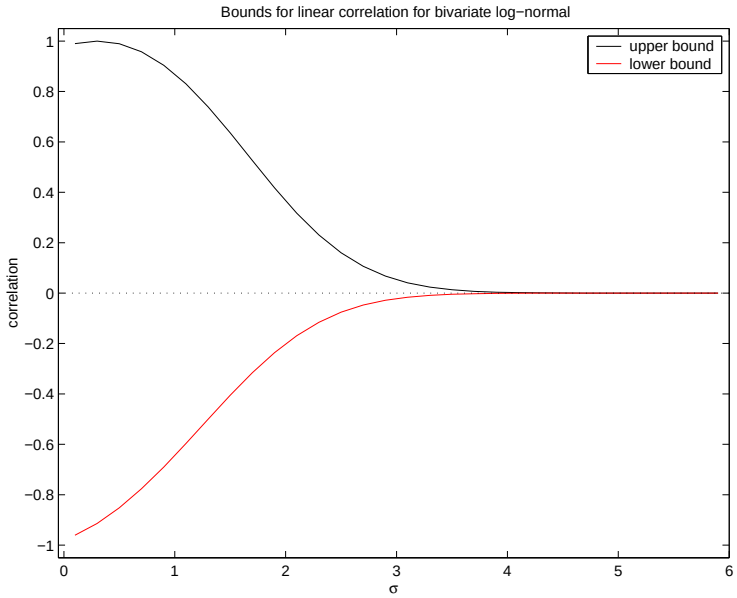


Figure 5.1: Bounds for linear correlation of bivariate log-normal

$$\rho_{\min} = \frac{\exp(-\sigma_X\sigma_Y) - 1}{\sqrt{(\exp(\sigma_X^2) - 1)(\exp(\sigma_Y^2) - 1)}} \leq 0$$

Now, we fix $\sigma_X = 0.3$, let σ_Y change. We plot the correlation between $\widetilde{X}$ and $\widetilde{Y}$ against σ_Y as shown in the Figure 5.1. We observe that

$$\lim_{\sigma_Y \to \infty} \rho_{\max} = \lim_{\sigma_Y \to \infty} \rho_{\min} = 0.$$

This example shows that we may have a random vector $(\widetilde{X}, \widetilde{Y})$, whose correlation is always zero, even when X and Y are perfectly dependent. This fact makes linear correlation an undesirable dependence indicator for the case.

5.3.2 Concordance Measures

As we discussed before, there are problems to use correlation coefficient as a measure of dependence. In this section, we'll discuss the modern concept of

concordance measure and introduce two popular concordance measures - Kendall's τ and Spearman's ρ.

Concordance is a concept for describing the association between random variables. Loosely speaking, a pair of random variables X and Y are concordant if large values of X tend to be associated with large values of Y, and similarly, small values of X tend to be associated with small values of Y.

Definition 5.14 *Let (x, y) and $(\tilde{x}, \tilde{y})$ be two independent observations from a vector (X, Y) of continuous random variables. Then, (x, y) and $(\tilde{x}, \tilde{y})$ are concordant if $(x - \tilde{x})(y - \tilde{y}) > 0$ and discordant if $(x - \tilde{x})(y - \tilde{y}) < 0$.*

The following theorem [47] reveals the relationship between concordance and copulas.

Theorem 5.15 *Let (X, Y) and $(\tilde{X}, \tilde{Y})$ be independent vectors of continuous random variables with joint distribution function H and $\tilde{H}$, respectively, with marginals F (of X and $\tilde{X}$) and G (of Y and $\tilde{Y}$). Let C and $\tilde{C}$ denote the copulas of (X, Y) and $(\tilde{X}, \tilde{Y})$, so that $H(x, y) = C(F(x), G(y))$ and $\tilde{H}(x, y) = \tilde{C}(F(x), G(y))$. Let Q denote the difference between the probability of concordance and discordance of (X, Y) and $(\tilde{X}, \tilde{Y})$:*

$$Q = \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) > 0\} - \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) < 0\}.$$

Then,

$$Q = Q(C, \tilde{C}) = 4 \int \int_{[0,1]^2} \tilde{C}(u, v) dC(u, v) - 1.$$

The Kendall's τ is defined in terms of concordance [35], the probability of concordance minus the probability of discordance.

Definition 5.16 *Let (X, Y) and $(\tilde{X}, \tilde{Y})$ be two independent observations, then the Kendall's τ is defined as:*

$$\tau = \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) > 0\} - \mathbb{P}\{(X - \tilde{X})(Y - \tilde{Y}) < 0\} \tag{5.11}$$

By applying Theorem 5.15, we have the following:

Proposition 5.17 *Let X and Y be continuous random variables whose copula is C. Then, the Kendall's τ for X and Y is given by:*

$$\tau = Q(C, C) = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1. \tag{5.12}$$

Note that the integral above is the expected value of the function $C(U, V)$ of uniform random variables U and V with joint distribution function C:

$$\tau = 4\mathbb{E}[C(U, V)] - 1.$$

To estimate τ from a sample of n pairs (X_i, Y_i), one notices that [24] :

$$\tau_{ij} = (+1)\mathbb{P}\{(X_i - X_j)(Y_i - Y_j) > 0\} + (-1)\mathbb{P}\{(X_i - X_j)(Y_i - Y_j) < 0\}$$

An unbiased estimator of Kendall's coefficient - Kendall's sample τ is given by:

$$\frac{2}{n(n-1)} \sum_{i<j}^{n} \operatorname{sgn}\left((X_i - X_j)(Y_i - Y_j)\right). \tag{5.13}$$

Spearman's ρ is another popular concordance measure. It's defined to be proportional to the probability of concordance minus the probability of the discordance :

Definition 5.18 *Let (X, Y), $(\tilde{X}, \tilde{Y})$, and (X', Y') be three independent observations, whose copula is C. Then, the Spearman's ρ is defined for two vectors (X, Y) and $(\tilde{X}, Y')$:*

$$\rho = 3(\mathbb{P}\{(X - \tilde{X})(Y - Y') > 0\} - \mathbb{P}\{(X - \tilde{X})(Y - Y') < 0\}). \tag{5.14}$$

Note that $\tilde{X}$ and Y' are independent, thus, the copula of $\tilde{X}$ and Y' is the product copula $\Pi(u, v) = uv$. Again, we apply Theorem 5.15 and have the following:

Proposition 5.19 *Let X and Y be continuous random variables whose copula is C. Then, the Spearman's ρ for X and Y is given by:*

$$\rho = 3Q(C, \Pi) = 12 \int \int_{[0,1]^2} uvdC(u, v) - 3. \tag{5.15}$$

The coefficient 3 in equation (5.14) is a normalization constant, since $Q(C, \Pi) \in [-1/3, 1/3]$. We also notice equation (5.15) can be rewritten as:

$$\begin{aligned} \rho &= 12 \int \int_{[0,1]^2} uvdC(u, v) - 3 \\ &= 12\mathbb{E}(UV) - 3 \\ &= \frac{\mathbb{E}(UV) - 1/4}{1/12} \\ &= \frac{\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)}{\sqrt{Var(U)}\sqrt{Var(V)}}, \end{aligned}$$

where the last equality comes from the fact that $1/2$ and $1/12$ are the mean and variance of standard uniform distribution. So, Spearman's ρ can be seen as the rank correlation.

The following Theorem [21] tells that copula based dependence measure Kendall's τ and Spearman's ρ will not have the problem for correlation coefficient.

Theorem 5.20 *Let X and Y be continuous random variables with copula C, and let κ denote Kendall's τ or Spearman's ρ. Then, the following are true:*

1. $\kappa(X,Y) = 1 \iff C(X,Y) = C^+$ *(comonotonic),*

2. $\kappa(X,Y) = 1 \iff C(X,Y) = C^-$ *(countermonotonic).*

We also notice that these two measures are increasing functions of the value of the copula. Thus, for continuous random variables, we can obtain all the values in the interval $[-1,1]$ for Kendall's τ or Spearman's ρ by a suitable choice of the copula. By Theorem 5.8, we also know that these two copula based measures are invariant under increasing transformation. In summary, Kendall's τ and Spearman's ρ are desirable alternate dependence measures.

5.3.3 Tail Dependence

Tail dependence is a concept that relates to the study of dependence between extreme values. It is used to describe the dependence in the upper-right-quadrant tail or lower-left-quadrant tail of a joint distribution. For example, we have the observation that the big price drop for one stock usually coincides with the slump of the stock price in the same sector.

Definition 5.21 *Let (X,Y) be a random vector with margins F and G. Let C be the copula. The lower tail dependence index is defined as:*

$$\lambda_L = \lim_{v \to 0^+} \mathbb{P}(Y \leq G^{-1}(v) | X \leq F^{-1}(v)),$$

given the limit λ_L exists. X and Y are said to be asymptotically dependent in the lower tail if $\lambda_L \in (0,1]$; X and Y are said to be asymptotically independent in the lower tail if $\lambda_L = 0$.

The lower tail dependence index λ_L is the limit of conditional probability that Y is less than the threshold value v, given that the X is, as v goes to 0. By applying the Bayes formula, we notice the following:

$$\begin{aligned}
\lambda_L &= \lim_{v\to 0^+} \mathbb{P}(Y \le G^{-1}(v) | X \le F^{-1}(v)) \\
&= \lim_{v\to 0^+} \frac{\mathbb{P}(Y \le G^{-1}(v), X \le F^{-1}(v))}{\mathbb{P}(X \le F^{-1}(v))} \\
&= \lim_{v\to 0^+} \frac{C(v,v)}{C(v,1)} \\
&= \lim_{v\to 0^+} \frac{C(v,v)}{v}.
\end{aligned}$$

We may also define the upper tail dependence through conditional probability.

Definition 5.22 *Let (X,Y) be a random vector with margins F and G. Let C be the copula. The upper tail dependence index is defined as:*

$$\lambda_U = \lim_{v\to 1^-} \mathbb{P}(Y > G^{-1}(v) | X > F^{-1}(v)),$$

given the limit λ_U exists. X and Y are said to be asymptotically dependent in the upper tail if $\lambda_U \in (0,1]$; X and Y are said to be asymptotically independent in the upper tail if $\lambda_U = 0$.

Notice $\mathbb{P}(Y > G^{-1}(v) | X > F^{-1}(v))$ can be written as:

$$\frac{1 - \mathbb{P}(X \le F^{-1}(v)) - \mathbb{P}(Y \le G^{-1}(v)) + \mathbb{P}(X \le F^{-1}(v), Y \le G^{-1}(v))}{1 - \mathbb{P}(X \le F^{-}(v))},$$

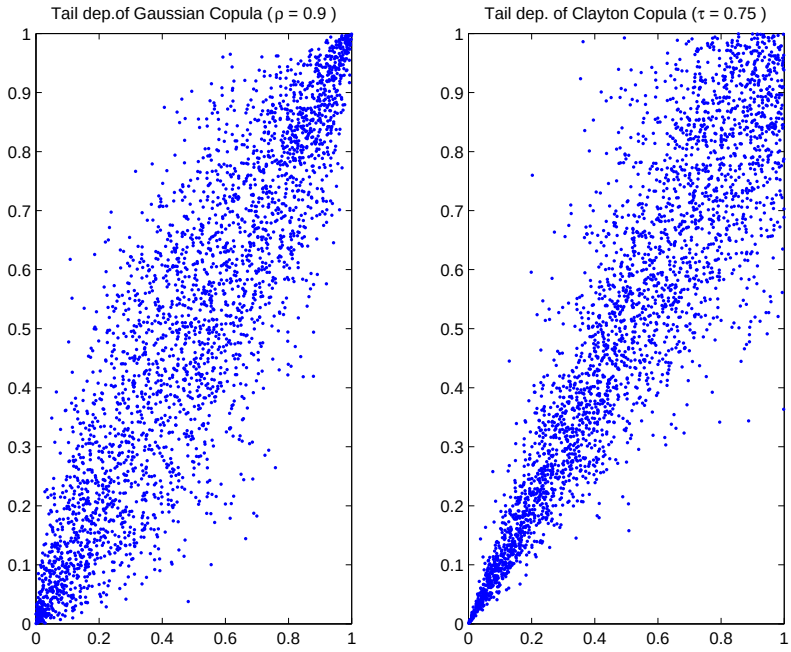


Figure 5.2: Tail dependence feature

we may write the upper tail dependence index as a function of copula as well.

$$\begin{aligned}
\lambda_U &= \lim_{v\to 1^-} \mathbb{P}(Y > G^{-1}(v) | X > F^{-1}(v)) \\
&= \lim_{v\to 1^-} \frac{\mathbb{P}(Y > G^{-1}(v), X > F^{-1}(v))}{\mathbb{P}(X > F^{-1}(v))} \\
&= \lim_{v\to 1^-} \frac{1 - C(v,1) - C(1,v) + C(v,v)}{C(v,1)} \\
&= \lim_{v\to 1^-} \frac{1 - 2v + C(v,v)}{1-v}.
\end{aligned}$$

The Figure 5.2 shows the tail dependence feature. The dependence structure are described by Gaussian (left) and Clayton (right) copulas respectively. The picture on the left shows no tail dependence and right figure shows the lower tail dependence. We will introduce Gaussian and Clayton in the following sections.

5.4 Gaussian Copulas

Gaussian copula was introduced by Li [38] for the pricing of basket credit derivatives. Its dependence structure is the same as multivariate Gaussian distribution. It belongs to the elliptical copulas family, where elliptical copulas are the copulas of elliptical distribution. We refer to Fang, Kotz and Ng [23] for details on elliptical distributions.

Definition 5.23 *The Gaussian copula is defined as follows:*

$$C_R^G(u_1, \cdots, u_n) = \Phi_R(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_n)), \tag{5.16}$$

where Φ_R is the multivariate Gaussian distribution function with zero mean and correlation matrix equal to R, Φ is the standard Gaussian distribution function, $u_i \in [0,1], i = 1, \cdots, n$.

We next discuss about the issue of simulation. Suppose we want to simulate a random vector $(v_1, \cdots, v_n)'$ with marginal distribution function $F_i, i = 1, \cdots, n$(VG distribution for our case). The dependence structure is described by Gaussian n-copula with correlation matrix R. In general, R is a symmetric strictly positive definite matrix. Thus, we have the Cholesky decomposition written as : $R = AA'$. Then, we may generate multivariate Gaussian random variable with dependence structure given by R. We next map the individual random variables to $[0,1]$ and finally, get the random vector with desired marginal by inverting the cumulative distribution function of specific marginal distribution.

1.	Find the Cholesky decomposition A of R.
2.	Simulate n independent random variates $z_1, \cdots, z_n$ from $N(0,1)$.
3.	Set $\boldsymbol{x} = A\boldsymbol{z}$.
4.	Set $u_i = \Phi(x_i), i = 1, \cdots, n$.
5.	$(v_1, \cdots, v_n)' = (F_1^{-1}(u_1), \cdots, F_n^{-1}(u_n))'$, F_i denotes the ith margin.

Table 5.1: Simulation Algorithm for Gaussian copula

We present the algorithm in Table 5.1. The Figure 5.3 shows how we can generate joint distribution by using Gaussian copula. The first row shows the density and level curve of Gaussian margins coupled by Gaussian copula, and the second row shows that of Student-t_3 margins coupled by Gaussian copula. The correlation coefficient $\rho = 0.8$ for both cases.

One factor Gaussian copula is another way to construct the Gaussian copula. This method is semi-explicit and we may obtain the multivariate distribution by numerical integration. We consider a Gaussian vector $(X_1, ..., X_n)$, let

$$X_i = \rho_i V + \sqrt{1 - \rho_i^2} W_i, i = 1, ..., n$$

where V, W_i are independent standard Gaussian random variables and $cov(X_i, X_j) = \rho_i \rho_j$, for $i \neq j$ and $cov(X_i, X_i) = 1$. Thus, the dependence is introduced through the single factor V, and the Gaussian copula is given by:

$$\begin{aligned} C(u_1, \cdots, u_n) &= \mathbb{P}(X_1 < \Phi^{-1}(u_1), \cdots, X_n < \Phi^{-1}(u_n)) \\ &= \mathbb{E}(\mathbb{P}(X_1 < \Phi^{-1}(u_1), \cdots, X_n < \Phi^{-1}(u_n)|V) \\ &= \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} \Phi \left(\frac{\Phi^{-1}(u_i) - \rho_i v}{\sqrt{1 - \rho_i^2}} \right) \right) \varphi(v) dv, \end{aligned}$$

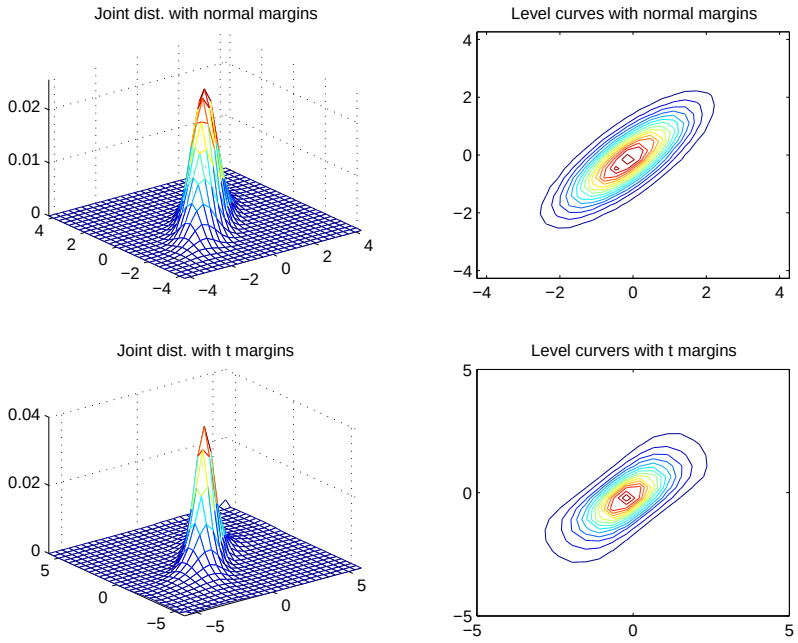


Figure 5.3: Density and level curves of Gaussian copula with different margins

where $\varphi(v)$ is the Gaussian density function. Then, the multivariate distribution of random vector $(v_1, \cdots, v_n)$ with margins F_i is given by:

$$\begin{aligned} F(v_1, \cdots, v_n) &= C(F_1(v_1), \cdots, F_n(v_n)) \\ &= \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} \Phi \left(\frac{\Phi^{-1}(F_i(v_i)) - \rho_i v}{\sqrt{1-\rho_i^2}} \right) \right) \varphi(v) dv. \end{aligned}$$

One may show that the upper and lower tail dependence index are equal to 0 for Gaussian copula [20]. This fact tells that Gaussian copula is not capable of describing dependence of extreme values and it builds dependence near the mean. These concerns lead us to consider other copulas being able to describe tail dependence.

5.5 Clayton Copulas

Clayton copula was introduced by Clayton [14]. It belongs to Archimedean copulas family. Detailed treatment of Archimedean copulas can be found in [47, 15]. Here, we follows [43] and introduce Clayton copula as a one factor model.

Let $G_i(x)$ be some distribution functions and let V be any positive random variable with density function $f(v)$. Then, we define the conditional distribution functions as conditionally independent given V and have the form:

$$\mathbb{P}(X_i < x|V = v) = G_i(x)^v.$$

Let $\psi(s)$ be the Laplace transform of V, then, the multivariate distribution will be:

$$\begin{aligned} F(x_1, \cdots, x_n) &= \mathbb{P}(X_1 < x_1, \cdots, X_n < x_n) \\ &= \mathbb{E}(\mathbb{P}(X_1 < x_1, \cdots, X_n < x_n|V) \\ &= \int_0^\infty \prod_{i=1}^n G_i(x_i)^v f(v)dv \\ &= \int_0^\infty \exp\left(v \sum_{i=1}^n \ln(G_i(x_i))\right) f(v)dv \\ &= \psi\left(-\sum_{i=1}^n \ln(G_i(x_i))\right). \end{aligned} \tag{5.17}$$

We may also get the implied unconditional marginal distribution $F_i(x_i)$ as:

$$\begin{aligned} F_i(x_i) &= \int_0^\infty G_i(x_i)^v f(v)dv \\ &= \int_0^\infty \exp(v \ln(G_i(x_i)))f(v)dv \\ &= \psi(-\ln(G_i(x_i))), \end{aligned}$$

from which we may solve $G_i(x_i)$:

$$G_i(x_i) = \exp(-\psi^{-1}(F_i(x_i))). \tag{5.18}$$

Plugging equation (5.18) into equation (5.17) and we get:

$$F(x_1, \cdots, x_n) = \psi\left(\sum_{i=1}^{n} \psi^{-1}(F_i(x_i))\right), \tag{5.19}$$

and the copula function is given by:

$$C(u_1, \cdots, u_n) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_n)). \tag{5.20}$$

So, copula C is an Archimedean copula with generator $\phi = \psi^{-1}$. If the positive random variable V follows Gamma distribution with parameter $1/\theta$, then we have

$$\begin{aligned} f(v) &= \frac{1}{\Gamma(1/\theta)} e^{-v} v^{1/\theta - 1} \\ \psi(s) &= (1+s)^{1/\theta} \\ \psi^{-1}(u) &= u^{-\theta} - 1 \end{aligned}$$

Thus, we have Clayton copula:

$$C(u_1, \cdots, u_n) = \left(\sum_{i=1}^{n} u_i^{-\theta} - n + 1\right)^{-1/\theta}.$$

We now discuss about the issue of simulation. Suppose we want to simulate a random vector $(x_1, \cdots, x_n)'$ with marginal distribution function $F_i, i = 1, \cdots, n$(VG distribution for our case). The dependence structure is defined by Clayton n-copula with parameter $\theta > 0$. We present the algorithm in Table 5.2.

One may show that for Clayton copula, the upper tail dependence index $\lambda_U = 0$ and the lower tail dependence index $\lambda_L = 2^{-1/\theta}$. So, Clayton copula has lower tail dependence as observed in Figure 5.2.

1.	Generate a r.v. $v \sim$Gamma$(1/\theta, 1)$, Laplace transform $\psi_v(s) = (1+s)^{-1/\theta}$.
2.	Generate independent uniform $(0,1)$ r.v.s $U_1, \cdots, U_n$.
3.	Calculate $Z_i = \psi\left(-\frac{1}{v}\ln(U_i)\right) = \left(1 - \frac{\ln(U_i)}{v}\right)^{-1/\theta}, i = 1, \cdots, n.$
4.	$(x_1, \cdots, x_n)' = (F_1^{-1}(Z_1), \cdots, F_n^{-1}(Z_n))'$.

Table 5.2: Simulation Algorithm for Clayton copula

5.6 Estimating Copula Models

In the statistical point of view, a copula function is a way to express a multivariate model. The Sklar's Theorem guarantees the existence of a unique copula function to a multivariate distribution given continuous margins. But, how to estimate the copula and margin is a subtle issue. One may estimate the copula parameters and marginal distributions simultaneously [15]. However, this method turns out to be very computationally expensive, since in the case of high dimension, the total number of parameters we need to calibrate will be very big. Joe and Xu [32] proposed that these set of parameters should be estimated in two steps, i.e., we first estimate the margins' parameters $\boldsymbol{\theta}_1$, then we estimate the copula parameters $\boldsymbol{\theta}_2$.

Now, we bring up another issue involved in implementing the Copula method. We recall that in Table 5.1 and 5.2, we need to evaluate the inverse of c.d.f. function, which is the distribution function of VG for our case. In practice, we calculate the table for $F(x)$ and find the $F^{-1}(y)$ by checking the table. We numerically calculate $F(x)$ by FFT and have the following result:

Proposition 5.24 *Let X be a random variable with characteristic function $\phi(x)$. Let $e^{\alpha x}$ be dampening factor. X has density function $f(x)$ and distribution function $F(x)$. Then, the c.d.f. function $F(x)$ is given by:*

$$F(x) = \frac{e^{\alpha x}}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{\phi(u + i\alpha)}{\alpha - iu} du \tag{5.21}$$

See Appendix for the proof. Thus, we may apply FFT method to estimate the distribution function.

5.7 Comparing with ICA Pricing

We next present an example of pricing the European-style basket option by two proposed methods. A basket option is an option whose payoff depends on the value of several underlying assets. The value of a call option is given by:

$$C(t, T, K) = e^{-r(T-t)} \mathbb{E} \left(\sum_{i=1}^{n} \alpha_i S_i(T) - K \right)^+ \tag{5.22}$$

We choose the stocks to be average weighted, which means the weights are chosen to be $\alpha_i = 1/n$. We are considering a basket with 10 underliers. By ticker symbol, the 10 stocks are: BA, C, GE, IBM, JNJ, KO, MRK, ORCL, PFE, and WMT. The strike prices are chosen to cover approximately 20% of the forward price of the basket on either side. For each asset, we simulate 100,000 paths. We first calculate the basket option price by ICA and we present the simulated prices in Table 5.3. In the table, for each strike maturity, we list the data for 3 different maturities. For each maturity, we list 3 items: the call price, the length of half of the 95% confidence interval, and the length relative to price.

	38 Days			367 Days			738 Days		
Strikes	Call	CIL	Percent	Call	CIL	Percent	Call	CIL	Percent
30.0000	13.6305	0.0174	0.0013	14.0285	0.0537	0.0038	14.6690	0.0734	0.0050
31.0000	12.6303	0.0174	0.0014	13.1124	0.0528	0.0040	13.8158	0.0723	0.0052
32.0000	11.6303	0.0174	0.0015	12.2133	0.0519	0.0042	12.9843	0.0712	0.0055
33.0000	10.6305	0.0174	0.0016	11.3331	0.0508	0.0045	12.1766	0.0699	0.0057
34.0000	9.6313	0.0174	0.0018	10.4744	0.0497	0.0047	11.3942	0.0686	0.0060
35.0000	8.6334	0.0173	0.0020	9.6405	0.0483	0.0050	10.6382	0.0671	0.0063
36.0000	7.6382	0.0172	0.0023	8.8325	0.0469	0.0053	9.9101	0.0656	0.0066
37.0000	6.6490	0.0170	0.0026	8.0532	0.0454	0.0056	9.2109	0.0639	0.0069
38.0000	5.6715	0.0167	0.0030	7.3050	0.0437	0.0060	8.5410	0.0622	0.0073
39.0000	4.7142	0.0162	0.0034	6.5908	0.0419	0.0064	7.9016	0.0604	0.0076
40.0000	3.7912	0.0154	0.0041	5.9121	0.0400	0.0068	7.2930	0.0585	0.0080
41.0000	2.9224	0.0143	0.0049	5.2705	0.0380	0.0072	6.7155	0.0566	0.0084
42.0000	2.1357	0.0128	0.0060	4.6673	0.0360	0.0077	6.1687	0.0546	0.0089
43.0000	1.4599	0.0109	0.0075	4.1044	0.0339	0.0083	5.6530	0.0526	0.0093
44.0000	0.9192	0.0089	0.0097	3.5822	0.0317	0.0089	5.1681	0.0506	0.0098
45.0000	0.5280	0.0068	0.0128	3.1020	0.0296	0.0095	4.7130	0.0485	0.0103
46.0000	0.2740	0.0049	0.0177	2.6632	0.0274	0.0103	4.2878	0.0465	0.0108
47.0000	0.1297	0.0033	0.0255	2.2666	0.0252	0.0111	3.8922	0.0444	0.0114
48.0000	0.0565	0.0022	0.0381	1.9112	0.0231	0.0121	3.5253	0.0424	0.0120
49.0000	0.0234	0.0014	0.0582	1.5965	0.0210	0.0131	3.1852	0.0403	0.0127
50.0000	0.0091	0.0008	0.0923	1.3202	0.0189	0.0143	2.8710	0.0383	0.0134

Table 5.3: Basket Option Prices for Different Maturities

It will be interesting for us to compare the risk-neutral moments of basket options and those of single names. We calibrate the risk-neutral VG measures for underlying assets and list the results in Table 5.4, 5.5 and 5.6.

Stocks	σ	ν	θ
BA	0.3372	0.0659	-0.4591
C	0.2977	0.0072	0.0853
GE	0.3510	0.1253	-0.3802
IBM	0.3476	0.0933	-0.6478
JNJ	0.3217	0.0631	-0.2579
KO	0.2569	0.0772	-0.3885
MRK	0.1720	0.0247	-1.3820
ORCL	0.5289	0.0039	0.4490
PFE	0.3034	0.0667	-0.2540
WMT	0.2844	0.1064	-0.3104

Table 5.4: Risk-neutral VG Parameters (38 Days)

We treat the basket of stocks as an asset and calibrate its risk-neutral VG measure as well. Then, we calculate the volatility, skewness and kurtosis from the VG parameters. We observe that the higher moments of basket option are in general less than the magnitudes of options on single name. We list the results in Table 5.7.

We also have a few comments on the skewness of risk-neutral distributions. It is observed [6] that individual risk-neutral distributions have less negative skewness than that of the market index. The reason is that the risk aversion investors are holding the index instead of holding the individual component stocks. Our results show that our basket option have smaller negative skewness. The explanation for the difference is that we are investigating an index composed of a very small pool of

Stocks	σ	ν	θ
BA	0.3225	0.2863	-0.4584
C	0.2117	0.0814	-1.1254
GE	0.2851	0.2786	-0.4942
IBM	0.2739	0.3807	-0.4694
JNJ	0.2225	0.4965	-0.3378
KO	0.1860	0.1944	-0.4884
MRK	0.2243	0.4905	-0.3044
ORCL	0.4911	0.4711	-0.5979
PFE	0.2853	0.0035	-0.1510
WMT	0.2440	0.4914	-0.3829

Table 5.5: Risk-neutral VG Parameters (367 Days)

Stocks	σ	ν	θ
BA	0.3340	0.5744	-0.3430
C	0.3481	0.4718	-0.2294
GE	0.3232	0.6784	-0.2507
IBM	0.3393	1.2670	-0.2465
JNJ	0.1642	0.8642	-0.3396
KO	0.2283	0.9016	-0.2102
MRK	0.2336	0.6093	-0.2544
ORCL	0.2419	0.2539	0.6994
PFE	0.2006	0.1957	0.3654
WMT	0.2588	1.2329	-0.2548

Table 5.6: Risk-neutral VG Parameters (738 Days)

assets. Its risk-neutral distribution doesn't reflect the strong risk aversion preference held by investors investing in very diversified portfolio like S&P.

	38 Days			367 Days			738 Days		
Stocks	Vol	Skew	Kurt	Vol	Skew	Kurt	Vol	Skew	Kurt
BA	0.3572	-0.2449	3.2384	0.4052	-0.8530	4.3731	0.4232	-1.2209	5.7781
C	0.2978	0.0062	3.0216	0.3846	-0.5486	3.4660	0.3821	-0.8016	4.8559
GE	0.3759	-0.3639	3.4661	0.3864	-0.9065	4.4240	0.3835	-1.2018	6.0441
IBM	0.4000	-0.4163	3.4001	0.3986	-1.1082	5.0296	0.4383	-1.8521	9.2370
JNJ	0.3282	-0.1468	3.2038	0.3258	-1.2695	5.6551	0.3558	-1.8251	8.0677
KO	0.2787	-0.3067	3.2959	0.2845	-0.8099	4.0599	0.3032	-1.6041	7.5407
MRK	0.2771	-0.2939	3.1372	0.3095	-1.2185	5.5368	0.3066	-1.3046	6.0398
ORCL	0.5296	0.0099	3.0118	0.6400	-1.1394	5.3366	0.4274	0.9639	4.4453
PFE	0.3104	-0.1613	3.2176	0.2854	-0.0056	3.0105	0.2576	0.7234	3.9584
WMT	0.3019	-0.3159	3.3870	0.3627	-1.2721	5.6466	0.3834	-2.0118	9.6298
Mean	0.3457	-0.2234	3.2379	0.3783	-0.9131	4.6538	0.3661	-1.0135	6.5597
Basket	0.2033	-0.1703	3.0957	0.2375	-0.9571	4.7707	0.2221	-0.7487	4.9350

Table 5.7: Risk-neutral Levels of Volatility, Skewness and Kurtosis for Single names and Basket Options

We next compare the Copula method with the ICA method. Essentially, copula based models are static models and are not good for modelling dependence through time. We show this by calibrating the copula parameters through ICA basket prices, and observe that the copula parameters are changing across different strike prices. And like the volatility smile and skew in the Black-Scholes model, we refer this fact to the skew and smile effect of parameters for one factor copula method. Thus, the one factor copula model will not do the job for pricing multiasset products for different strikes and maturities.

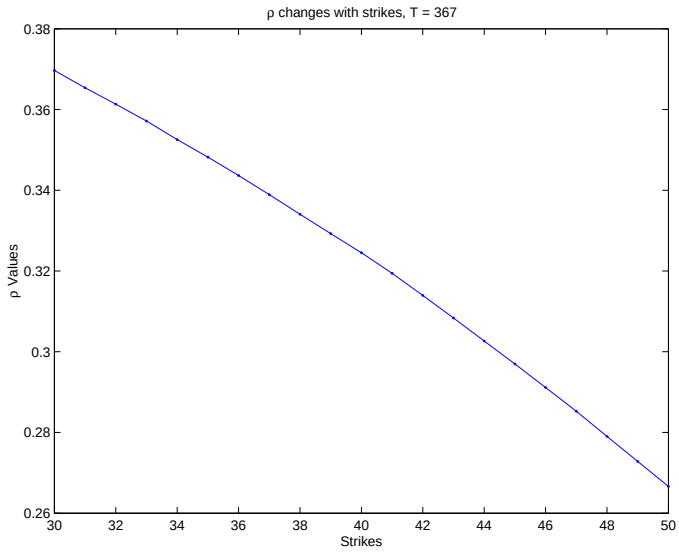


Figure 5.4: Gaussian Copula Correlation Skew (367 Days)

As of implementation, we adopt the point of view proposed by Joe and Xu [32], i.e., we estimate the copula model in two steps. We first estimate the risk-neutral measure for each asset, the results are listed in Table 5.4, 5.5 and 5.6.

We next calibrate the copula parameters. We are considering two copulas: one factor Gaussian copula and one factor Clayton copula. We calibrated the parameters for both methods and found them inconsistent across different strikes and maturities. For Gaussian copula, Figure 5.4, 5.5 show the changing correlation across different strikes. We put these two figures together in Figure 5.6 showing that the correlation parameters are inconsistent across different maturities.

For Clayton copula, Figure 5.7, 5.8 and 5.9 show the changing dependence parameters across different strikes. We put these three figures together in Figure 5.10 showing that the dependence parameters are inconsistent across different maturities.

We also do the statistical analysis for the copula parameters and present the

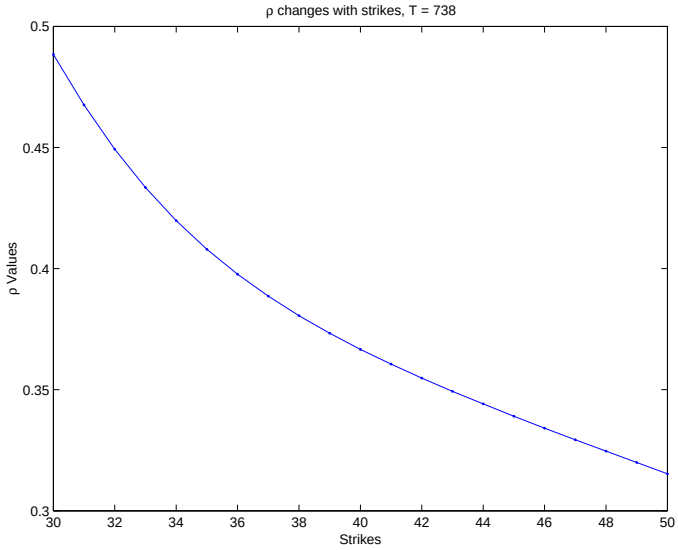


Figure 5.5: Gaussian Copula Correlation Skew (738 Days)

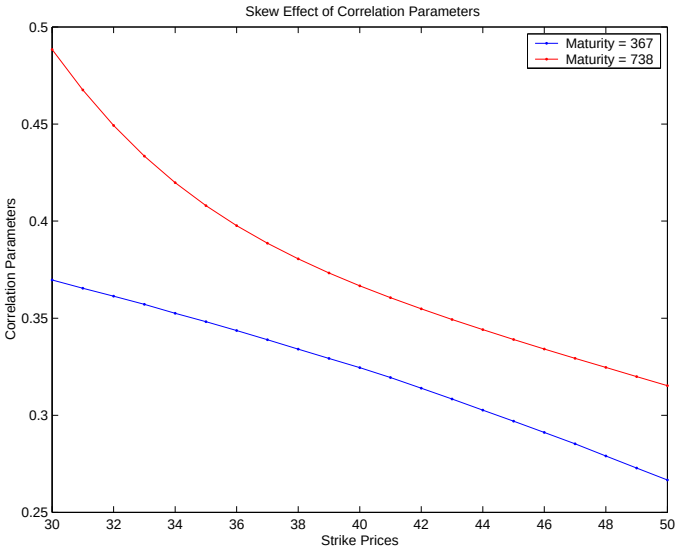


Figure 5.6: Gaussian Copula Correlation Skew

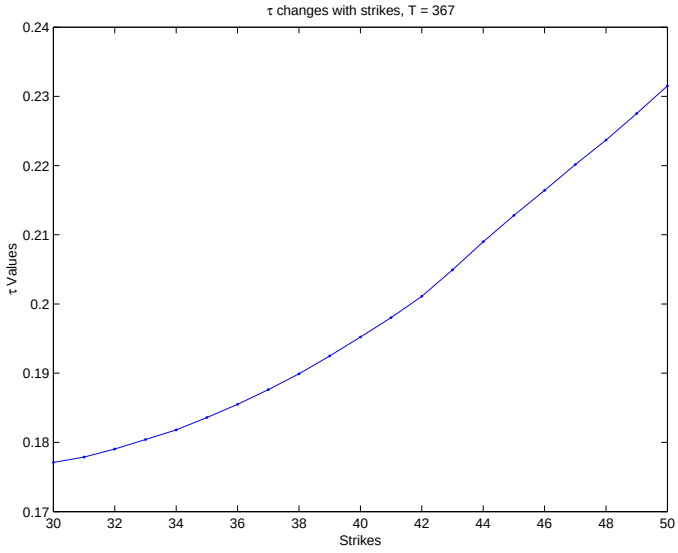


Figure 5.7: Clayton Copula Parameter Skew & Smile (367 Days)

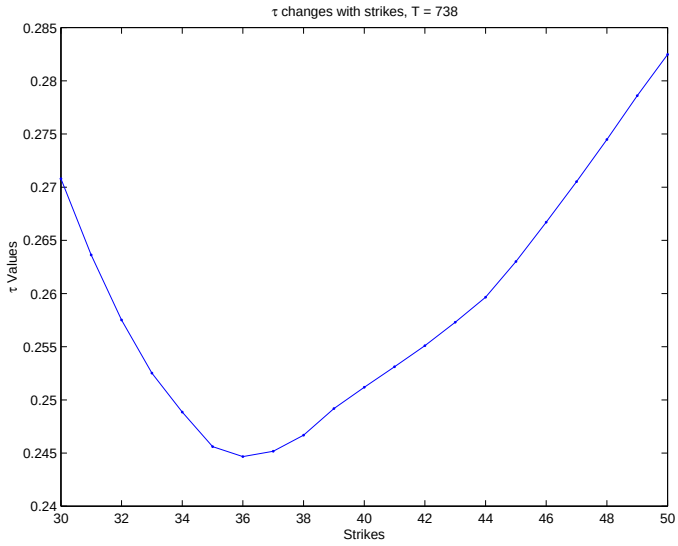


Figure 5.8: Clayton Copula Parameter Skew & Smile (738 Days)

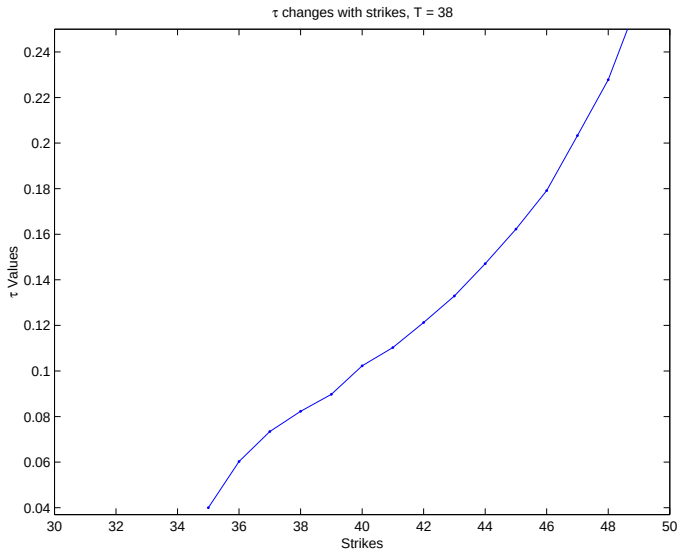


Figure 5.9: Clayton Copula Parameter Skew & Smile (38 Days)

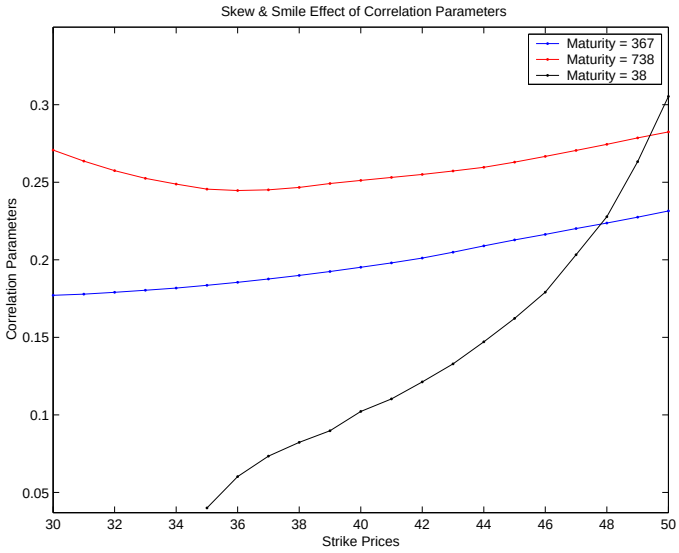


Figure 5.10: Clayton Copula Parameter Skew & Smile

result in Table 5.8.

	Gaussian		Clayton	
Maturities	Mean	STD	Mean	STD
367 Days	0.3220	0.0319	0.1988	0.0176
738 Days	0.3783	0.0502	0.2589	0.0115

Table 5.8: Statistical Analysis for Copula Parameters

We observed that the standard deviation of parameters from Clayton method is less than that from Gaussian method, which tells that one factor Clayton copula is more consistent with different strike prices than Gaussian copula. This is not surprising since Clayton copulas have lower tail dependence while Gaussian copulas have no tail dependence at all.

In general, the copula method is a static method. It builds the dependence through random variable and the dependence structure does not evolve through time. There are documented skew in applying copula method to CDO prices [4]. In our numerical example, we observe the skew and smile as well.

In conclusion, we explored how to extend VG processes to multiasset product pricing by ICA. We assume that the log-return of each asset is driven by independent VG factors. We uncover the factors by applying ICA and estimate the physical measures of factors by MLE. We then apply the transformation of measure to get the risk-neutral measures. We derive the log price of each asset in terms of linear sum of risk-neutral VG factors. Thus, we may apply the liquid option market prices

to calibrate the measure change parameters. Then, we apply the measure change parameters to derive the risk-neutral measure. Finally, we price the multiasset products by doing Monte Carlo simulation. Future study will focus on how to apply this methodology to more general Lévy processes.

Appendix A

Correlation Between Two Log-normal Random Variables

Proposition A.1 *Let X_1 and X_2 be two normal random variables with parameters (μ_1, σ_1),(μ_2, σ_2). Let ρ be the correlation between X_1 and X_2. Let $Y_1 = \exp(X_1)$ and $Y_2 = \exp(X_2)$. Then, the correlation between them:*

$$\rho(Y_1, Y_2) = \frac{\exp(\rho\sigma_1\sigma_2) - 1}{\sqrt{(\exp(\sigma_1^2) - 1)(\exp(\sigma_2^2) - 1)}}. \tag{A.1}$$

Proof. Note that $\mathbb{E}(Y_1Y_2) = \mathbb{E}\left(\exp(X_1 + X_2)\right) = \mathbb{E}\left(\exp(X)\right)$. where $X \sim N(\mu_1 + \mu_2, \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)$. Thus, $\exp(X)$ follows the log-normal distribution. So, we have

$$\mathbb{E}(Y_1Y_2) = \exp(\mu_1 + \mu_2 + \frac{1}{2}(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2))$$

Then,

$$\begin{aligned} Cov(Y_1, Y_2) &= \mathbb{E}(Y_1Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) \\ &= \exp\left(\mu_1 + \mu_2 + \frac{1}{2}(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)\right) - \exp\left(\mu_1 + \frac{1}{2}\sigma_1^2\right)\exp\left(\mu_2 + \frac{1}{2}\sigma_2^2\right) \\ &= \exp\left(\mu_1 + \mu_2 + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)\right)(\exp(\rho\sigma_1\sigma_2) - 1). \end{aligned}$$

We also know the variance of log-normal distribution:

$$\begin{aligned} Var(Y_1) &= \exp(2\mu_1 + \sigma_1^2)(\exp(\sigma_1^2) - 1), \\ Var(Y_2) &= \exp(2\mu_2 + \sigma_2^2)(\exp(\sigma_2^2) - 1). \end{aligned}$$

Finally, we apply the linear correlation formula and get

$$\begin{aligned}\rho(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)}\sqrt{Var(Y_2)}} \\ &= \frac{\exp(\rho\sigma_1\sigma_2) - 1}{\sqrt{(\exp(\sigma_1^2) - 1)(\exp(\sigma_2^2) - 1)}},\end{aligned}$$

where $\rho = \rho(X_1, X_2)$.■

Appendix B

Numerical Approximation for the Distribution Function

Proposition B.1 *Let X be a random variable with characteristic function $\phi(x)$ and let $e^{-\alpha x}$ be the dampening factor. X has density function $f(x)$ and distribution function $F(x)$. Then, the c.d.f. function $F(x)$ is given by:*

$$F(x) = \frac{e^{\alpha x}}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\frac{\phi(u + i\alpha)}{\alpha - iu}du \tag{B.1}$$

Proof. The characteristic function for X is given by:

$$\phi(u) = E\exp(iuX) = \int_{-\infty}^{\infty} e^{iux}f(x)dx$$

Note the Fourier transform of $e^{-\alpha x}F(x)$ is given by :

$$\begin{aligned}\int_{-\infty}^{\infty} e^{iux}e^{-\alpha x}F(x)dx &= \int_{-\infty}^{\infty} e^{(iu-\alpha)x}\left(\int_{-\infty}^{x} f(y)dy\right)dx \\ &= \int_{-\infty}^{\infty}\left(\int_{y}^{\infty} e^{-(\alpha-iu)x}f(y)dx\right)dy \\ &= \int_{-\infty}^{\infty} f(y)\frac{e^{-(\alpha-iu)y}}{\alpha - iu}dy \\ &= \frac{1}{\alpha - iu}\int_{-\infty}^{\infty} f(y)e^{iy(u+\alpha i)}dy \\ &= \frac{\phi(u + i\alpha)}{\alpha - iu}\end{aligned}$$

Then, by taking the inverse Fourier transform, the distribution function $F(x)$ is given by:

$$F(x) = \frac{e^{\alpha x}}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{\phi(u + i\alpha)}{\alpha - iu} du$$

This completes the proof. ■

BIBLIOGRAPHY

[1] Aldrich, J. (1997), "R.A. Fisher and the making of maximum likelihood 1912-1922," *Statistical Science*, 12, No. 3 , 162-176.

[2] Alexander, C. (2001), *Market Models*, John Wiley & Sons, Inc., New York.

[3] Bates, D. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutschemark Options, " *Review of Financial Studies*, 9, 69-108.

[4] Berd, A., R. Engle and A. Voronov (2005), "The Underlying Dynamics of Credit Correlations," *Preprint.*

[5] Burtschell, X., J. Gregory and J.-P. Laurent (2005), "A Comparative analysis of CDO Pricing Models," *Working Paper*

[6] Bakshi, G., N. Kapadia and D. Madan (2003), "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 16, No. 1, 101-143.

[7] Barndorff-Nielsen, O.E. (1995), "Normal Inverse Gaussian Distributions and the Modeling of Stock Returns," Research Reprot no. 300, Department of Theoretical Statistics, Aarhus University.

[8] Barndorff-Nielsen, O.E. and N. Shephard (2000), "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics," *Journal of the Royal Statistical Society*, Series B 63, 167-241.

[9] Carr, P., H. Geman, D. Madan and M. Yor (2001), "The Fine Structure of Asset Returns: An Emperical Investigation," *Journal of Business*, 75, 2, 305-332.

[10] Carr, P., H. Geman, D. Madan and M. Yor (2003), "Stochastic Volatility for Lévy Processes," *Mathematical Finance*, Vol. 13, No. 3, 345-382.

[11] Carr, P., H. Geman, D. Madan and M. Yor (2005), "Self Decomposability and Option Pricing," *Preprint*.

[12] Chourdakis, K. (2004), "Option Pricing Using the Fractional FFT," *Preprint*.

[13] Cherbini,U. and E. Luciano,(2002), "Bivariate Option Pricing with Copula," *Applied Mathematical Finance*, 8, 69-85.

[14] Clayton, D. G. (1978), "A model for association in bivariate life talbes and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika* 65, 141-151.

[15] Cherubini, U., E. Luciano, and W. Vecchiato (2004), *Copula Methods in Finance*, Wiley Finance.

[16] Carr, P. and D. B. Madan (1999), "Option Valuation using the Fast Fourier Transform," *Journal of Computational Finance*, 2, 61-73.

[17] Cover, T.M. and J. A. Thomas (1991), *Elements of Information Theory*, John Wiley & Sons Inc., New York.

[18] Cont, R. and P. Tankov (2004), *Financial Modelling With Jump Processes*, Chapman & Hall/CRC.

[19] Eberlein, E., U. Keller and K. Prause (1998), "New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model," *Journal of Business*, Vol. 71, No. 3, 371-405.

[20] Embrechts, P., F. Lindskog, A. McNeil (2003), "Modelling Dependence with Copulas and Applications to Risk Management," In:*Rachev, S. (Eds.)Handbook of Heavy Tailed Distributions in Finance*, Elsevier, Chapter 8, pp. 329-384.

[21] Embrechts, P., A. McNeil, D. Straumann(2002), "Correlation and dependence in risk management: properties and pitfalls," In: *Dempster, M. et al. (Eds.), Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, pp. 176- 223.

[22] Fama, E. F. (1965), "The Behavior of Stock-Market Prices," *Journal of Business*, Vol. 38, No. 1, 34-105.

[23] Fang, K.T., S. Kotz, and K.W. Ng (1987), *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[24] Gibbons, J.D. (1992), *Nonparametric Statistical Inference*, Marcel Dekker, New York.

[25] Goorbergh, R.W.J. van den, C. Genest and B.J.M. Werker(2003), "Multivariate Option Pricing Using Dynamic Copula Models," *Working paper.*

[26] Hyvärinen, A (1999), "Survey on Independent Component Analysis," *Neural Computing Surveys*, Vol. 2, pp. 94-128.

[27] Hyvärinen, A (1999), "Fast and Robust Fixed-point Algorithms for Independent Component Analysis," *IEEE Trans. on Neural Networks*, 10, 626-634, 1999.

[28] Hull, J.C. (2002), *Options Futures and Other Derivatives (Fifth Edition)*, Prentice Hall.

[29] Hyvärinen, A, J. Kaerhunen and E. Oja (2001), *Independent Component Analysis*, John Wiley & Sons Inc., New York.

[30] Harvey, C. R., and A. Siddique (2000), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263-1295.

[31] Joe, H. (1997), *Multivariate Models and Dependence Concepts*,Chapman & Hall, New York.

[32] Joe, H. and J.J. Xu (1996), "The Estimation Method of Inference Functions for Margins for Multivariate Models," Technical Report No. 166, Department of Statistics, University of British Columbia.

[33] Jolliffe, I.T. (1986), *Principal Component Analysis*, Springer-Verlag.

[34] Jacod, J. and A. Shiryaev (2003), *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin.

[35] Kendall, M.G. (1938), "A new measure of rank correlation," *Biometrika*, 30, 81-93.

[36] Kraus, A., and R. Litzenberger (1976), "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, 31, 1085-1100.

[37] Laurent, J.-P. and J. Gregory (2002), "Basket Default Swaps, CDO's and Factor Copulas," *Working Paper.*

[38] Li, D. X. (2000), "On Default Correlation: A Copula Function Approach," *Journal of Fixed Income*, 9, 43-54.

[39] Ling, C. H. (1965), "Representation of associative functions," *Publicationes Mathematicae, Debrecen*, 12, 189-212.

[40] Luciano, E and W. Schoutens (2005), "A Multivariate Jump-Driven Financial Asset Model," *Working Paper.*

[41] Markowitz, H. M. (1952), "Portfolio Selection," *Journal of Finance*, 7, 77-91.

[42] Madan, D. B., P. Carr, and E. Chang (1998), "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2, 79-105.

[43] Madan, D. B., M. Konikov and M. Marinescu (2004), "Credit and Basket Default Swaps," *Working Paper.*

[44] Madan, D. B., F. Milne (1991), "Option Pricing with VG Martingale Components," *Mathematical Finance*, Vol.1, No.4, 39-55.

[45] Madan, D. B. and E. Seneta (1990), "The Variance Gamma (V.G.) Model For Share Market Returns," *Journal of Business*, Vol.63, No.4,511-524.

[46] Madan, D. B. and J. Yen (2004), "Asset Allocation for CARA Utility with Multivariate Lévy Returns," *Preprint.*

[47] Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer-Verlag, New York.

[48] Oksendal, B. (1995), *Stochastic Differential Equations: An introduction with Applications (Fourth Edition)*, Springer-Verlag Berlin Heidelberg.

[49] Sato, K.(1999), "Lévy Processes and Infinitely Divisible Distributions," Cambridge University Press.

[50] Scott, L. (1997), "Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Application of Fourier Inversion Methods," *Mathematical Finance*, 7, 413-426.

[51] Schoutens, W. (2002), "Meixner Processes: Theory and Applications in Finance," *EURANDOM Report*, 2002-004.

[52] Schoutens, W. (2003), *Lévy Processes in Finance: Pricing Financial Derivatives*, Wiley & Sons Inc.

[53] Schweizer, B. and A. Sklar (1961), "Associative functions and statistical triangle inequalities," *Publicationes Mathematicae, Debrecen*, 8, 169-186.

[54] Schweizer, B. and A. Sklar (1983), *Probabilistic Metric Spaces*, North-Holland, New York.

[55] Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," *Publications de I'Institut de Statistique de L'Université de Paris*, 8, 229-231.

[56] Sklar, A. (1996), "Random variables, distribution functions, and copulas - a personal look backward and forward," in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendorff, B. Schweizer, and M. Taylor, pp. 1-14, Institute of Mathematical Statistics, Hayward, CA.

[57] Wang, S.S. (1998), "Aggregation of Correlatied Risk Portfolios: Models - Algorithms," CAS Committee on Theory of Risk.